

Humboldt-Universität zu Berlin
Philosophische Fakultät I
Institut für Bibliotheks- und Informationswissenschaft

Datenmodellierung für digitale Editionen –
Stand und Perspektiven zwischen XML/TEI, Linked Open Data
und Ontologien

Masterarbeit im Rahmen des Weiterbildenden Masterstudiengangs Bibliotheks- und
Informationswissenschaft im Fernstudium

vorgelegt von
Till Meyer

Gutachter:

Prof. Vivien Petras, PhD

Dr. Christian Stein

Leipzig, Mai 2019

Verteidigt am 18.09.2019

Inhaltsverzeichnis

1. Einleitung	1
2. Semantic Web: Grundlagen und Verbreitung im Kulturerbebereich	13
3. Semantic-Web-Technologien & digitale Editionen – Theoretische Ansätze	25
4. Semantic-Web-Technologien & digitale Editionen in der Praxis	42
5. Fazit	51
Literaturverzeichnis	53
Anhang	63

1. Einleitung

Im Jahr 2001 formulierte Tim Berners-Lee zum ersten Mal die Idee eines Semantic Web¹, also eines umfassenden digitalen Informationsraums, der für Maschinen nicht nur lesbar, sondern ‚verständlich‘ ist und das überkommene, ‚grobmaschige‘ und in seiner vollen Bedeutung nur Menschen zugängliche Web of Documents ergänzen sollte. Auf diese Initialzündung folgten weitere Anregungen, die das Konzept des Semantic Web aufgriffen. Mit dem W3-Konsortium existierte schon früh eine Organisation, die die Entwicklung des Semantic Web, insbesondere durch Ausformulierung grundlegender Standards wie RDF, OWL und SPARQL, gezielt vorantrieb, und schon 2002 fand die erste International Semantic Web Conference statt. Seitdem wurde das Konzept des Semantic Web in Form unterschiedlicher, in ihrer Bedeutung teils deckungsgleicher, teils variierender Schlagwörter wie Giant Global Graph, Web 3.0, Web of Data oder Linked (Open) Data in vielen Bereichen aufgegriffen. Von der freien Wirtschaft über mehr oder weniger alle Domänen der Wissenschaft bis hin zu Kulturerbeinstitutionen wurden seine vielversprechenden Anwendungsmöglichkeiten erörtert, erprobt und teilweise bereits umgesetzt. Das Ergebnis ist allein schon quantitativ beeindruckend: Die Linked Open Data Cloud ist mittlerweile auf über 1200 Datensets angewachsen² – damit erreicht sie eine Größe, die in herkömmlicher Form schon gar nicht mehr grafisch darstellbar ist, weswegen Visualisierungen der Linked Open Data Cloud oft auf alte, überschaubarere Versionen zurückgreifen müssen.³ In Form des Google Knowledge Graph oder von Amazons Open Graph protocol haben semantische Technologien, freilich unter Verzicht auf den Aspekt der freien (Nach-)Nutzbarkeit, schon vor einiger Zeit auch Eingang in weitverbreitete kommerzielle Anwendungen gefunden.

Mit einer gewissen Verzögerung, dafür in den letzten Jahren mit umso größerem Interesse, sind Ansätze aus dem Semantic-Web-Bereich auch in der Bibliothekswelt aufgegriffen und diskutiert worden. Dies verdeutlicht schon ein Blick auf einen beliebten Indikator für die Entwicklungen im deutschsprachigen Bibliothekswesen: die Deutschen Bibliothekskarte bzw. die dort gehaltenen Vorträge. Wer die auf dem Opus-Publikationsserver abgeleg-

¹ Vgl. Tim Berners-Lee, James Handler, Ora Lassila: The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, in: Scientific American 284 (2001), Issue 5, S. 34–43; dt. als: Mein Computer versteht mich. In: Spektrum der Wissenschaft, August 2001, S. 42–49.

² Vgl. <https://lod-cloud.net/> [23.04.2019]

³ So beispielsweise im Artikel „Linked Open Data“ der deutschsprachigen Wikipedia, der den Stand vom September 2011 abbildet; vgl. https://de.wikipedia.org/wiki/Linked_Open_Data [24.04.2019]

ten Vortragsfolien auf entsprechende Schlagworte hin untersucht, bekommt recht schnell einen Eindruck davon, welche Bedeutung dem Thema mittlerweile auch im Bibliothekssektor zugemessen wird. Wer etwa das Stichwort „Semantic Web“ eingibt, erhält den ersten Treffer für das Jahr 2005 – bezeichnenderweise handelt es sich um einen Vortrag über das „Imageproblem der deutschen Bibliothekswissenschaft“, der die „Entwicklung von Informationstechnologie (bspw. Semantic Web, Ontologien)“ als eines der zentralen Forschungs- und damit wohl auch Zukunftsfelder bibliothekarischer Arbeit herausstellt⁴ –, im folgenden Jahr fehlen Vorträge zum Thema, 2007 sind wieder drei, 2008 zwölf Vorträge zu verzeichnen; bis zum jüngsten, 108. Bibliothekartag sind mit einer Ausnahme (im Jahr 2016 gab es nur zwei Vorträge zum Thema) immer mindestens fünf, in manchen Jahren elf (2017), zwölf (2010 und 2011) oder 13 (2012) Vortragsfolien unter dem Stichwort „Semantic Web“ auffindbar.⁵ Eine ähnliches Ergebnis erhält man bei der Suche nach dem Stichwort „Linked Open Data“: Nach einer ersten Erwähnung im Jahr 2009 mehren sich die Treffer rasch von fünf (2010) auf acht (2011) und pendeln sich, mit einigen Ausschlägen nach oben (im Jahr 2017 behandeln ganze 27 Vorträge das Thema) bei zehn bis 15 ein; insgesamt sind seit 2009 140 Vorträge unter dem Stichwort verzeichnet!⁶

Ein gänzlich anderes Bild bietet der Blick auf einen Bereich, der dem Bibliothekswesen gewissermaßen benachbart und mit ihm in vielerlei Hinsicht, v.a. institutionell und inhaltlich, verbunden ist: das (digitale) Editionsweisen. Wer etwa die Vorträge der Jahrestagung des DHd-Verbundes, einer der wohl wichtigsten Plattformen für Fragen des digitalen Editionswezens im deutschsprachigen Raum, nach Stichwörtern wie „Semantic Web“ im Zusammenhang mit Editionen absucht, erkennt schnell, dass dieser Thematik hier bei Weitem nicht die Bedeutung zugemessen wird wie auf den Deutschen Bibliothekartagen. Es ist daher auch kein Zufall, dass das Standardwerk zum Thema, Patrick Sahles dreibändige

⁴ Vgl. Petra Hauke et al.: Das Imageproblem der deutschen Bibliothekswissenschaft. Vortragsfolien für den 94. Deutschen Bibliothekartag 2005. <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/90>.

⁵ https://opus4.kobv.de/opus4-bib-info/solrsearch/index/search/start/0/rows/20/sortfield/score/sortorder/desc/searchtype/simple/query/semantic+web/facetNumber_year/all/doctypenfq/Konferenzfolien [23.04.2019]

⁶ <https://opus4.kobv.de/opus4-bib-info/solrsearch/index/search/start/0/rows/20/sortfield/score/sortorder/desc/searchtype/simple/query/Linked+Open+Data/doctypenfq/Konferenzfolien> [23.04.2019]

[!] Dissertation über „Digitale Editionsformen“⁷, das Thema nicht behandelt. Dies ist umso erstaunlicher, als eines der großen Versprechen des Semantic Web darin besteht, bisher implizites Wissen nicht nur maschinenlesbar zu explizieren, sondern auf dieser Grundlage auch über verschiedenste strukturelle und inhaltliche Grenzen hinweg auszutauschen und synergetisch zueinander in Beziehung zu setzen. Mit diesem Potenzial, heterogene Informationen zu erkennen und zu vernetzen, bietet das Semantic Web auch einen Ansatz, eines der zentralen Probleme in den digitalen Geisteswissenschaften im Allgemeinen und im digitalen Editionswesen im Besonderen im großen Maßstab zu überwinden: das Problem der ‚Datensilos‘.⁸ Auf XML-Basis zeichnen sich – trotz oder gerade wegen beständig anwachsender Datenbestände und einer kaum zu überschauenden Anzahl und Vielfalt von Editionsprojekten – für dieses Problem bis dato zwar Verbesserungsansätze,⁹ aber keine grundstürzenden Lösungen ab.

An dieser Stelle will diese Arbeit ansetzen. Ihre zentrale Frage lautet: Welchen Mehrwert können Semantic-Web-Technologien für digitale Editionen bieten? Dieser Thematik soll im Folgenden weniger auf einer technischen, als auf einer ‚konzeptuellen‘ Ebene nachgegangen werden. Mit ‚konzeptuell‘ ist an dieser Stelle eine Herangehensweise gemeint, die ihren Fokus auf die Evaluation von auf Semantic-Web-Technologien beruhenden Konzepten zur digitalen inhaltlichen Erschließung von Texten legt. Die entsprechenden Ansätze sollen dabei unter zwei verschiedenen Blickwinkeln, respektive Frageebenen, analysiert werden. Die erste Ebene bezieht sich auf die von der Fachgemeinschaft erörterten Potentiale, die Semantic-Web-basierte Ansätze für digitale Editionen *theoretisch* bieten. Die erkenntnisleitende Frage lautet hier in erster Linie: Wie könnten formale Wis-

⁷ Vgl. Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels [= Schriften des Instituts für Dokumentologie und Editorik 7-9], 3 Bände, Norderstedt 2013. Man mag einwenden, dass Sahles Arbeit bereits vor sechs Jahren veröffentlicht wurde. Allerdings wurden bereits zu diesem Zeitpunkt Semantic-Web-Technologien in vielen Kulturerbebereichen eingesetzt oder in ihren Anwendungsmöglichkeiten diskutiert; die Deutsche Nationalbibliothek beispielsweise stellte schon seit 2010 ihre Normdaten als Linked Open Data im RDF-Format zur freien Verfügung; vgl. Abschnitt 2 dieser Arbeit.

⁸ Vgl. Torsten Schrade: Geisteswissenschaftliche Fachdatenrepositorien im Semantic Web. Modellierung, Vernetzung, Visualisierung. Conference paper and presentation at the 3rd DHd conference in Leipzig, 2016, S. 232–235. <http://www.dhd2016.de/abstracts/vorträge-054.html>.

⁹ Ein Beispiel ist der von der Berlin-Brandenburgischen Akademie der Wissenschaften getragene Webservice *correspsearch*, in dessen Rahmen „Verzeichnisse verschiedener digitaler und gedruckter Briefeditionen nach Absender, Empfänger, Schreibort und Datum“ aggregiert und damit zentral durchsuchbar gemacht werden. Damit ist *correspsearch* sicherlich ein nützliches Forschungs-Werkzeug; es bleibt aber, unter anderem durch seine Begrenzung auf Metadaten von Briefeditionen, weit hinter den (potenziellen) Möglichkeiten eines Semantic-Web-Ansatzes zurück; vgl. <https://correspsearch.net/index.xql?l=de> (dort auch Zitat).

sensmodelle wie Ontologien die Inhaltserschließung verbessern? An diese zentrale Frage sind weitere Problemstellungen angelagert, beispielsweise: Lassen sich Konzepte wie Linked Open Data, der Giant Global Graph usw. für die Arbeit mit digitalisierten Texten nutzbar machen? Welche Synergien verspricht der Rückgriff auf bereits existierende Semantic-Web-Infrastrukturen, insbesondere die Datenbestände der Bibliotheken und anderer ‚natürlicher strategischer Partner‘ des digitalen Editions wesens aus dem Kulturerbebereich? Auf der zweiten Ebene nähere ich mich dem Untersuchungsgegenstand des Mehrwerts von Semantic-Web-Technologien für digitale Editionen unter einem gänzlich anderen Vorzeichen; dieser Abschnitt stellt gewissermaßen das praktisch-empirische Gegenstück zur theoretisch-explorativen Herangehensweise der ersten Ebene dar. Hier soll in erster Linie in Form einer quantitativen Analyse der Datenmodelle digitaler Editionen die aktuelle Editions*praxis* in den Blick genommen werden, um die Frage zu beantworten, welche Szenarien des Einsatzes von Semantic-Web-Technologien momentan überhaupt umgesetzt werden.

Diese Fragestellung wurde auf der Basis einer Matrix operationalisiert (vgl. Abb. 1). Die Grundlage dieser Matrix stellen zwei Dimensionen dar, die ich mit den Begriffen *Abstraktion* und *Expressivität* bezeichnen möchte. Der Abstraktionsgrad gibt an, wieweit Editionskonzepte von der Möglichkeit des Semantic Web Gebrauch machen, nicht nur konkrete Entitäten (Menschen, Dinge usw.), sondern auch abstrakte Konzepte als Informationsressourcen zu erfassen und zu verarbeiten. Ich unterscheide innerhalb dieser Dimension drei Stufen: Der ersten Stufe werden – unabhängig vom Charakter des edierten Materials – alle Projekte zugeordnet, die nur konkrete Entitäten erfassen, also in der Regel Personen, Orte und Institutionen. Den Stufen 2 und 3 werden Projekte zugeordnet, die über die bloße Erschließung von im Text genannten konkreten Entitäten hinausgehen, indem sie – meist in Form von Schlagwörtern, die abstrakte Konzepte repräsentieren – auch tieferliegende Informationsschichten erfassen. Der Unterschied zwischen der zweiten und der dritten Stufe ergibt sich aus dem edierten Material: Hoch strukturiertes Quellenmaterial, das einen seriellen oder semi-seriellen Charakter aufweist wird in Stufe 2 eingruppiert – dies betrifft v.a. Rechnungs- und Verwaltungsschriftgut wie Zollregister, Jahrrechnungen, Universitätsmatrikel, Baumeister- oder Urfehdebücher. Komplexeres Material, das über keine derartige ausgeprägte Strukturen verfügt, wird Stufe 3 zugeordnet – hierzu zählen u.a. Briefe, Tagebücher, Reisebeschreibungen und Vertragswerke. Die zweite Dimension bezeichnet den Grad der semantischen Expressivität, mit der die Entitäten zueinander in Beziehung gesetzt werden. Auch hier unterscheide ich zwischen drei Graden bzw. Stufen: Der ersten,

	Expressivitätsgrad 1: Normdaten zur Referenzierung	Expressivitätsgrad 2: Normdaten als LOD	Expressivitätsgrad 3: Modellierung interner Bezüge
Abstraktionsgrad 1: konkrete Entitäten, stark und schwach strukturierte Texte	1	2	3
Abstraktionsgrad 2: konkrete und abstrakte Entitäten, stark strukturierte Texte	4	5	6
Abstraktionsgrad 3: konkrete und abstrakte Entitäten, schwach strukturierte Texte	7	8	9

Abb. 1: Abstraktions-/Expressivitätsmatrix (schematische Darstellung)

einfachsten Stufe ordne ich Editionsprojekte zu, die zwar auf prinzipiell Semantic-Web-fähige Datensätze, insbesondere Normdatensätze der GND, zurückgreifen, diese aber nur für eine eindeutige Referenzierung ihrer Ressourcen nutzen, ohne dabei die Möglichkeiten des Semantic Web auszuschöpfen. Erst wenn im Sinne des Semantic Web eine Vernetzung der Ressourcen der Edition mit anderen Ressourcen vorliegt, ist die zweite Stufe der semantischen Expressivität erreicht. Die dritte Stufe ist schließlich dadurch definiert, dass die Entitäten der jeweiligen Edition nicht nur – wie in Stufe 2 – mit Daten der LOD-Cloud, sondern auch untereinander in Form eines formalen Modells verbunden werden und so in einen qualitativen Zusammenhang treten.

Aus der Kombination der beiden Dimensionen *Abstraktion* und *Expressivität* ergibt sich ein Raster, auf das sich Modellierungskonzepte im Hinblick auf ihre Verwendung von Semantic-Web-Technologien einordnen lassen. Die Matrixfelder 1 und 9 markieren die äußersten Pole: Das Konzept des Matrixfelds 1 beschränkt sich darauf, konkrete Entitäten lediglich zum Zwecke der Referenzierung mit einem Normdatensatz zu verbinden, während Ma-

trixfeld 9 das elaboriertesten Anwendungskonzept von Semantic-Web-Technologien darstellt, indem es eine hohe semantische Expressivität mit einer tiefen Erschließung kombiniert und auf die komplexesten Quellenmaterialien anwendet. An dieser Grundidee, orientiert an einer konzeptuellen Matrix aus Abstraktions- und Expressivitätsgrad die theoretischen Potenziale von Semantic-Web-Technologien für digitale Editionen mit der tatsächlichen Editionspraxis abzugleichen, orientiert sich im Grunde auch der Aufbau dieser Arbeit. Allerdings ist den entsprechenden Ausführungen noch ein einführender Komplex (Abschnitt 2) vorgeschaltet, in dem einige meines Erachtens für die Untersuchung notwendige Kontexte erörtert werden. Dafür werde ich zum einen die Grundidee des Semantic Web und seine technischen Grundlagen zu umreißen versuchen, zum anderen synoptisch darstellen, welche darauf aufbauenden Infrastrukturen, insbesondere im Kulturerbebereich, sich in den letzten Jahren entwickelt haben und als natürliche Anknüpfungspunkte für digitale Editionen dienen könnten.

Wie bereits skizziert, werde ich im folgenden Teil die verschiedenen theoretischen Ansätze zur Inhaltserschließung im Rahmen von digitalen Editionen auf Basis von Semantic-Web-Technologien in den Mittelpunkt meiner Darstellung rücken (Abschnitt 3). Auch hier ist dem eigentlichen Hauptteil ein kleinerer kontextualisierender Abschnitt vorangestellt. Dieser beschäftigt sich mit den tradierten Standards der Datenmodellierung, also insbesondere mit der Text Encoding Initiative (TEI), und der an ihnen geübten Kritik. Es liegt in der Natur der Sache, dass ein Kapitel über die theoretischen Potenziale von Semantic-Web-Technologien fast ausschließlich auf der Auswertung der einschlägigen Fachliteratur beruht. Einen vollständigen Überblick zu geben ist angesichts der Fülle an Texten zum Thema Semantic Web kaum möglich, weswegen ich mich an dieser Stelle auf einen kurzen Abriss derjenigen Literatur beschränke, die genuin an der Schnittstelle von digitalem Editonswesen und Semantic Web operiert, während ich den großen Bereich der thematisch mittelbar relevanten Literatur, insbesondere aus dem Kulturerbebereich¹⁰, an dieser Stelle aussparen muss.

Thematisch einschlägige Beiträge sind in vor allem in zwei Zusammenhängen entstanden. Zum ersten sind hier die Texte zu nennen, die sich der Frage, ob und wie sich Semantic-Web-Technologien nutzen lassen, um Defizite des digital-editorischen Quasi-Standards TEI zu beheben, in erster Linie theoretisch nähern. Schon 2006 fassten beispielsweise

¹⁰ Vgl. z.B. als jüngeres Beispiel Andreas Vlachidis et al.: Semantic Representation and Enrichment of Cultural Heritage Information for Fostering Reinterpretation and Reflection on the European History, in: Marinos Ioannides (Hg.), Digital Cultural Heritage, Cham 2018, S. 91–103.

Tumarello et al. die Möglichkeit ins Auge, mithilfe von RDF und OWL das Problem des überlappenden Markups zu überwinden.¹¹ Dieser Ansatz wurde in den letzten Jahren verschiedentlich aufgegriffen¹² und – namentlich von Andreas Kuczera – weiterentwickelt.¹³ Spätestens seit dem Release der jüngsten TEI-Version P5 und den damit einhergehenden Erweiterungen der Datenmodellierung wurde zudem unter verschiedenen Prämissen der Versuch unternommen, die semantische Expressivität des TEI-Markups durch diverse Ontologie-Erweiterungen zu erhöhen.¹⁴ In eine entsprechende Richtung zielte bereits 2009 der Beitrag von Ore et al., der die Möglichkeiten des Mappings zentraler Elemente von TEI auf das CIDOC CRM zum Gegenstand hatte.¹⁵ Eine ähnliche Stoßrichtung, nur noch stärker auf die Entwicklung einer eigenen ‚TEI-Ontologie‘ gerichtet, verfolgten vor Kurzem

¹¹ Vgl. Giovanni Tumarello et al.: A novel Textual Encoding paradigm based on Semantic Web tools and semantics, in: Proceedings of the 5th Edition of the International Conference on Language Resources and Evaluation, Paris 2006, S. 247–252. http://www.lrec-conf.org/proceedings/lrec2006/pdf/225_pdf.pdf.

¹² Vgl. als Übersicht Thomas Efer: Introducing NoXML for the Digital Humanities, in: Proceedings of INFORMATIK 2017, Lecture Notes in Informatics (LNI), hg. von der Gesellschaft für Informatik, 2017, S. 1161–1666. DOI:10.18420/in2017_117.

¹³ Vgl. Andreas Kuczera: Digital Editions beyond XML – Graph-based Digital Editions, in: Histoinformatics 2016 – The 3rd Histoinformatics Workshop. Proceedings of the 3rd Histoinformatics Workshop on Computational History (Histoinformatics 2016), hrsg. von Marten Düring, Adam Jantowt, Johannes Preiser-Kappeller, co-located with Digital Humanities 2016 conference (DH 2016), Krakow, 2016, S. 37–46. http://ceur-ws.org/Vol-1632/paper_5.pdf; ders.: Graphentechnologien in den Digitalen Geisteswissenschaften, in: ABI Technik 37 (2017), Heft 3, S. 179–196. DOI: 10.1515/abitech-2017-0042. Jüngst widmete die *Zeitschrift für digitale Geisteswissenschaften* der Thematik Graphdatenbanken für geisteswissenschaftliche Editionen und andere Forschungsvorhaben einen Sonderband (vgl. Andreas Kuczera, Thorsten Wübbena, Thomas Kollatz (Hgg.): Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten [= Zeitschrift für Ideengeschichte, Sonderband 4], Wolfenbüttel 2019. DOI: 10.17175/sb004.). Dieser von Kuczera und anderen verfolgte Ansatz, dem seit 2017 sogar jährliche Tagungsreihen an der Akademie der Wissenschaften zu Mainz gewidmet sind, stützt sich zwar auf graphbasierte Technologien, allerdings wird dabei meist auf die Verwendung von URIs, RDF oder vergleichbaren Standards verzichtet. Damit handelt es sich hier streng genommen zwar um eine Graph-, aber keine Semantic-Web-Technologie, weswegen die entsprechenden Ansätze in den folgenden Ausführungen auch keine weitere Beachtung finden.

¹⁴ Für einen älteren, aber instruktiven Überblick vgl. Anna Jordanous, Alan Stanley, Charlotte Tupman: Contemporary transformation of ancient documents for recording and retrieving maximum information: when one form of markup is not enough, in: *Proceedings of Balisage: The Markup Conference 2012*. Balisage Series on Markup Technologies 8 (2012). <https://doi.org/10.4242/BalisageVol8.Jordanous01>.

¹⁵ Vgl. Christian-Emil Ore, Øyvind Eide: TEI and cultural heritage ontologies: Exchange of information?, in: *Literary and Linguistic Computing*, Vol. 24, No. 2, 2009, S. 161–172. <https://doi.org/10.1093/lc/fqp010>.

Ciotti et al.,¹⁶ die zuvor schon entsprechende Überlegungen im Zusammenhang mit dem Projekt *Geolat* angestellt hatten.¹⁷ Auf keine spezielle Ontologie, sondern generell auf die verschiedenen Ansätze der Verbindung von TEI-Markup mit externen Ontologien und deren jeweilige Implikationen gerichtet, ist darüber hinaus der Text von Øyvind Eide über „Ontologies, Data Modeling, and TEI“.¹⁸

Weitere, weniger theoretische Untersuchungen entstanden im Zusammenhang mit ersten explorativen Projekten zum Einsatz von Semantic-Web-Technologien für digitale Editionen. Zu nennen wären hier u.a. die Beschreibungen der *Henry III Fine Rolls*,¹⁹ des Projekts *Semantic Blumenbach*²⁰, der *Digitalen Edition der Baseler Urfehdebücher*²¹ oder die auf der Bookkeeping Ontology aufbauenden Projekte.²² Von zentraler Bedeutung für die vorliegende Untersuchung sind darüber hinaus einige Arbeiten, die sich der Thematik auf einer eher theoretisch-konzeptionellen Ebene nähern und die daher in Abschnitt 3 ausführlicher ausgewertet werden sollen. Neben den überblicksartigen Beiträgen von Jörg

¹⁶ Vgl. Fabio Ciotti, Francesca Tomasi: Formal Ontologies, Linked Data, and TEI Semantics, in: Journal of the Text Encoding Initiative 9 (2016/17). DOI: 10.4000/jtei.1480.

¹⁷ Vgl. Fabio Ciotti, Maurizio Lana, Francesca Tomasi: TEI, Ontologies, Linked Open Data. Geolat and Beyond, in: Journal of the Text Encoding Initiative 8 (2014 /15). DOI: 10.4000/jtei.1365.

¹⁸ Vgl. Øyvind Eide: Ontologies, Data Modeling, and TEI, in: Journal of the Text Encoding Initiative 8 (2014/15). DOI: 10.4000/jtei.1191.

¹⁹ Vgl. Arianna Ciula, Jose Miguel Monteiro Viera: Implementing an RDF/OWL Ontology on Henry the III Fine Rolls, in: OWL. Experiences and Directions Third International Workshop 2007. <http://www.cidoc-crm.org/Irmoo/Resources/implementing-an-rdfowl-ontology-on-henry-the-iii-fine-rolls>; dies., Paul Spence: Expressing complex associations in medieval historical documents – The Henry III Fine Rolls Project, in: Literary and Linguistic Computing. The journal of digital scholarship in the humanities 23 (2008), No. 3, S. 311–325. <https://doi.org/10.1093/lc/fqn018>.

²⁰ Vgl. Jörg Wettlaufer et al.: Semantic Blumenbach. Exploration of Text–Object Relationships with Semantic Web Technology in the History of Science, in: Digital Scholarship in the Humanities 30 (2015), Issue suppl. 1, S. 187–198. <https://doi.org/10.1093/lc/fqv047>.

²¹ Vgl. Christopher Pollin, Georg Vogeler: Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the "Urfehdebücher der Stadt Basel", in: Proceedings of the Second Workshop on Humanities in the Semantic Web, co-located with 16th International Semantic Web Conference (DH 2017), Wien, 2017, S. 27–32. <http://ceur-ws.org/Vol-2014/paper-03.pdf>.

²² Vgl. Georg Vogeler: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? in: Constanze Baum, Thomas Stäcker (Hgg.), Grenzen und Möglichkeiten der Digital Humanities [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 1]. Wolfenbüttel 2015. DOI: 10.17175/sb001_007; ders.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, in: Jürgen Sarnowsky (Hg.), Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters, Göttingen 2016, S. 13–42. DOI: 10.14220/9783737006774.13.

Wettläufer²³ und Roland S. Kamzelak²⁴ sowie den kodikologisch ausgerichteten Ausführungen von Robert Kummer²⁵ ist hier vor allem Stefan Dumonts Aufsatz „Briefe kommentieren im Semantic Web – ein Konzept“²⁶ zu nennen sowie Jochen Vogelers Versuch der Definition einer „assertive edition“, die die Möglichkeit der Informationsextraktion aus historischen Texten auf Basis von RDF-Statements erörtert.²⁷ Ebenfalls für den theoretischen Rahmen der Thematik wichtig, und daher noch Gegenstand ausführlicherer Untersuchung, sind die beiden Texte von Jochen Strobel über „Digitale Briefedition und semantische Erschließung“²⁸ und von Stefan Heßbrüggen-Walter über eine graphbasierte Analyse philosophischer Diskurse.²⁹

Wie bereits erwähnt, soll dem eben skizzierten theoretischen Teil ein empirisch-praktischer folgen, in dessen Mittelpunkt die Analyse der Datenmodelle digitaler Editionen steht (Abschnitt 4). Zu diesem Zweck habe ich ein Korpus jüngerer Editionsprojekte erstellt. Die von mir getroffene Auswahl beschränkt sich von vornherein auf bestimmte Typen von Editionen, und zwar solche, für die eine Anwendung von Semantic-Web-Technologien zur Inhaltserschließung am ehesten gewinnversprechend erscheint. Konkret schließt dies vor allem Editionen von Ego-Dokumenten, also Briefe, Tagebücher, Reisebeschreibungen und

²³ Vgl. Jörg Wettlaufer: Der nächste Schritt? Semantic Web und digitale Editionen, in: Roland S. Kamzelak, Timo Steyer (Hgg.), *Digitale Metamorphose: Digital Humanities und Editionswissenschaft* [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 2]. Wolfenbüttel 2018. DOI: 10.17175/sb03.

²⁴ Vgl. Roland S. Kamzelak: Chancen und Grenzen von Normdaten, FRBR und RDF, in: „Ei, dem alten Herrn zoll' ich Achtung gern“, Festschrift für Joachim Veit zum 60. Geburtstag, München 2016, S. 423–435. DOI: 10.25366/2018.29.

²⁵ Vgl. Robert Kummer: Semantic Technologies for Manuscript Descriptions — Concepts and Visions, in: Franz Fischer, Christiane Fritze, Georg Vogeler (Hgg.), *Kodikologie und Paläographie im digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2* [= Schriften des Instituts für Dokumentologie und Editorik 3], Norderstedt 2011, S. 133–154. <http://kups.ub.uni-koeln.de/id/eprint/4347>.

²⁶ Vgl. Stefan Dumont: Briefe kommentieren im Semantic Web – ein Konzept, in: DARIAH-DE Working Papers Nr. 33. Göttingen 2019. urn:nbn:de:gbv:7-dariah-2019-5-8.

²⁷ Vgl. Georg Vogeler: The ‘assertive edition’ : On the consequences of digital methods in scholarly editing for historians, in: *International Journal of Digital Humanities* 1 (2018). <http://dx.doi.org/10.17613/M6JS9H76P>.

²⁸ Vgl. Jochen Strobel: Digitale Briefedition und semantische Erschließung. Von den Briefen der Jenaer Romantikergeneration zur Edition der Korrespondenz August Wilhelm Schlegels, in: *editio* 28 (2014), S. 151–174.

²⁹ Vgl. Stefan Heßbrüggen-Walter: What People Said: The Theoretical Foundations of a Minimal Doxographical Ontology and Its Use in the History of Philosophy, in: Constanze Baum, Thomas Stäcker (Hgg.), *Grenzen und Möglichkeiten der Digital Humanities* [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 1]. Wolfenbüttel 2015. DOI: 10.17175/sb001_001.

dergleichen ein. Ebenfalls Aufnahme in das Korpus haben darüber hinaus Editionen von Verwaltungs-, Rechnungs- und vergleichbarem historisch relevanten Schriftgut (Sitzungsprotokolle, Urfehdebücher, Matrikelbücher) gefunden. Nicht oder kaum³⁰ berücksichtigt wurden hingegen Projekte, die deutlich literaturwissenschaftlich ausgerichtet sind, da bei diesen das Hauptaugenmerk der Edition weniger auf der Erschließung historischer Informationen als auf Fragen der Textvarianz, respektive –genese, liegt.³¹ Darüber hinaus wurden Editionsprojekte nicht berücksichtigt, deren Ziel die Erschließung fremdsprachlichen Materials ist oder deren Projektbeginn zu weit zurückliegt, um ernsthaft den Einsatz von Semantic-Web-Technologien in Erwägung zu ziehen. Als Stichjahr wurde hier das Jahr 2008 festgelegt. Letztlich fiel die Wahl auf dieses Datum, weil in diesem Jahr die meines Wissens weltweit erste Edition auf Semantic-Web-Basis, die bereits erwähnten *Fine Rolls Henry III*, der Öffentlichkeit vorgestellt und damit die hier zu erörternden Frage- und Problemstellungen erstmals in Ansätzen praktisch erprobt wurden. Diese Festlegung ist freilich nicht sakrosankt, andere Stichjahre oder zeitliche Eingrenzungen, etwa die Beschränkung auf aktuell laufende Projekte, wären denkbar gewesen. Letzteres wurde bewusst verworfen, sodass neben aktuellen auch bereits abgeschlossene Projekte Eingang in das Untersuchungskorpus gefunden haben. Mit anderen Worten: Ein beispielsweise von 2010 bis 2015 laufendes Editionsprojekt würde berücksichtigt werden, ein 2006 begonnenes und noch laufendes Projekt nicht.

Nach diesen Kriterien habe ich versucht, ein Korpus vollständig oder teilweise (hybrid) digitaler Editionen zu erstellen. Es gibt bekanntermaßen keine offizielle zentrale ‚Verzeichnisstelle‘ für (digitale) Editionen, weswegen es nicht leicht ist, sich einen Überblick über laufende und abgeschlossene Projekte in diesem Bereich zu verschaffen. Als eine der ausführlichsten Übersichten zum Thema kann sicherlich der von Patrick Sahle gepflegte „Catalog of Digital Scholarly Editions“ gelten, der nach aktuellem Stand (Februar 2019) immerhin 461 Einträge verzeichnet.³² Von den für diese Arbeit infrage kommenden 104 Editionen von deutschsprachigem Material konnte allerdings ein Großteil nicht in das Untersuchungskorpus aufgenommen werden, da die betreffenden Projekte entweder vor

³⁰ Ausnahmen stellen Projekte wie *Stefan George Digital* oder die *Uwe Johnson-Werkausgabe* dar, die entweder aufgrund ihres besonderen Datenmodells (*George*) oder ihres ‚Mischcharakters‘ (die *Johnson-Werkausgabe* enthält neben literarischen Texten auch Briefe) Eingang in das Korpus gefunden haben.

³¹ Zum Unterschied zwischen historisch-kulturwissenschaftlichem und philologisch-literaturwissenschaftlichem Interesse an Texten und den daraus resultierenden divergenten Erwartungen an deren digitale Edition vgl. Vogeler, ‚Assertive Edition‘, S. [1]–[3].

³² Vgl. <http://www.digitale-edition.de/> [27.04.2019]

2008 ins Leben gerufen worden oder zu stark literaturwissenschaftlich ausgerichtet waren. Die verbliebenen gut 30 Editionen wurden mit einem weiteren digitalen Katalog, dem „Catalogue Digital Editions“, abgeglichen und nach dem erwähnten Schema – digitale Editionen deutschsprachigen Materials ab 2008 – um rund 10 Projekte ergänzt.³³ Nach den gleichen Kriterien habe ich aus den Webseiten häufig an Editionsprojekten beteiligter Institutionen (Wissenschaftsakademien³⁴, Herzog August Bibliothek Wolfenbüttel³⁵, Literaturarchiv Marbach³⁶) weitere Projekte gefiltert; davon wurden (nach ‚Abzug‘ der bereits in den beiden digitalen Katalogen genannten Editionen) rund 20 in das Korpus aufgenommen. Eine weitere Quelle war die Datenbank GEPRIS, die alle von der DFG in irgendeiner Form geförderten Projekte, also auch Editionen, verzeichnet.³⁷ Insgesamt konnte so eine Liste von 77 für diese Arbeit relevanten Editionen erstellt werden.

Bringt man die oben erläuterten Auswahlkriterien (deutschsprachig; frühestens 2008 in Angriff genommen; eher geschichts- und kultur- denn literaturwissenschaftlich orientiert) in Rechnung, bildet das so gewonnene Korpus sicherlich nicht das digitale Editions-wesen in seiner Gesamtheit ab. Dennoch dürfte es unabhängig von seiner Gewichtung aufgrund der Auswahlkriterien einen repräsentativen Querschnitt zumindest für gewisse Typen von digitalen Editionen darstellen.

Um das Korpus im Zuge dieser Untersuchung analysieren zu können, habe ich alle Projekte daraufhin untersucht, welche Arten von Entitäten erfasst (Abstraktionsgrad) und ob dabei Normdaten wie die GND verwendet werden, welches Datenmodell zur Anwendung kommt sowie ob und in welcher Form dabei auf Semantic-Web-Standards zurückgegriffen wird. Darüber hinaus habe ich weitere kontextualisierende Informationen erfasst, etwa zur Projektlaufzeit oder zu den am Projekt beteiligten Institutionen. Zum einen habe ich dafür die Internetseiten der jeweiligen Projekte nach relevanten Informationen abgesucht. Da die Projektbeschreibungen allerdings nicht immer eindeutig oder ausführlich genug sind, habe

³³ Vgl. <https://dig-ed-cat.acdh.oeaw.ac.at/> [27.04.2019]

³⁴ Da die zentrale Webseite der deutschen Akademien zu ihren Forschungsvorhaben (https://www.akademienunion.de/forschungsprojekte/?no_cache=1) nicht ganz vollständig sind, musste ich die entsprechenden Projektwebseiten der jeweiligen Akademien in Berlin (BBAW), Göttingen, Heidelberg, Leipzig, Mainz und München sowie der Österreichischen Akademie der Wissenschaften in Wien einzeln untersuchen.

³⁵ <http://www.hab.de/de/home/bibliothek/digitale-bibliothek-wdb/digitale-editionen/liste-aller-digitalen-editionen.html> [27.04.2019]

³⁶ <https://www.dla-marbach.de/digital-humanities/projekte/> [27.04.2019]

³⁷ <https://gepris.dfg.de/gepris/OCTOPUS.jsessionid=813C8093B23C03FBD5FA77901355C7C6> [27.04.2019]

ich mich entschlossen, die oben genannten Punkte in Form eines Fragebogens zusammenzufassen und an die jeweiligen Projektverantwortlichen zu senden. Von den insgesamt 77 Projekten erhielt ich 39 Rücksendungen, von denen 38 einen ausgefüllten Fragebogen enthielten; ein Projektverantwortlicher lehnte die „Veröffentlichung der Daten unserer Edition“ ab. Die restlichen 38 Anfragen blieben unbeantwortet. Für 14 von ihnen erwiesen sich letzten Endes die Editionsbeschreibungen auf den einschlägigen Projektwebseiten oder die Sekundärliteratur als aussagekräftig genug, um alle für die Matrix relevanten Informationen zusammenzutragen. Die verbliebenen 24, für die weder ein Fragebogen noch eine ausführliche Editionsbeschreibung vorlag, konnten in dieser Arbeit nicht analysiert werden, sodass sich ein Gesamtkorpus von insgesamt 52 Editionen ergibt, auf das sich in dieser Arbeit zurückgreifen konnte.³⁸

Wie bereits erwähnt, ist die Auswertung dieses kombinierten, aus den Projektbeschreibungen und den Fragebögen gewonnen Korpus zentraler Gegenstand von Abschnitt 4, in dem ich die 52 Editionen der bereits erwähnten Abstraktions-Komplexitäts-Matrix zuordne. An dieser Stelle werde ich aber auch ausgewählte Datenmodelle näher beleuchten, wenn sie für eines der Matrixfelder paradigmatisch erschienen.

In einem letzten Schritt (Abschnitt 5) werde ich in einer Zusammenfassung die theoretische und die empirisch-praktische Perspektive auf die Frage, welchen Mehrwert Semantic-Web-Technologien für digitale Editionen bieten können, abschließend miteinander abgleichen.

³⁸ Vgl. für eine Gesamtübersicht die Tabelle im Anhang.

2. Semantic Web: Grundlagen und Verbreitung im Kulturerbereich

Die Vision des Semantic Web wurde erstmals 2001 in einem mittlerweile berühmt gewordenen Aufsatz mit dem programmatischen Titel „The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities“ umrissen.¹ In diesem Beitrag adressierten die Autoren, unter ihnen mit Tim Berners-Lee eine der Gründerfiguren des World Wide Web, die Defizite des damaligen Internets. Der Kern der Kritik richtete sich gegen eines seiner Grundprinzipien, nämlich den Umstand, dass es als eine Verknüpfung von Dokumenten, also Webseiten, konzipiert worden war. Dokumente sind immer schon aus mehreren Informationen zusammengesetzt, die wiederum zu Informationen in anderen Dokumenten eine spezifische Beziehung besitzen können. Während aber der menschliche Nutzer durch sein Kontextwissen in die Lage versetzt wird, den semantischen Gehalt der jeweiligen Informationen zu ermitteln und zwischen ihnen Bezüge herzustellen, bleibt diese in natürlicher Sprache formulierte Informationsgranularität Maschinen unzugänglich. Für sie stellt das Web 1.0 eine Ansammlung von unstrukturiert in Dokumenten aufeinanderfolgenden Daten dar, deren Verbindung mittels eines Hyperlinks weder spezifische Informationen über die Art ihrer Verbindung noch ihre Bedeutung enthalten. Um diese Diskrepanz zu überwinden, entwickelte Berners-Lee die Vision eines Semantic Web als „extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.“² Diese Definition enthält bereits alle maßgeblichen Eckpfeiler des Semantic Web, nämlich:

- 1.) Das bestehende Web soll nicht ersetzt, sondern von einem Web of Documents zu einem Web of Data erweitert oder ‚vertieft‘ werden;
- 2.) Informationen sollen mit Aussagen über ihre Bedeutung verknüpft und somit semantisch ‚aufgeladen‘ werden, und zwar auf eine Weise, die von Maschinen prozessierbar ist;
- 3.) Ziel des Semantic Web ist eine Verbesserung der Mensch-Maschine-Kommunikation. Um dies zu ermöglichen, wurde ein neuer Ansatz zur Strukturierung von Informationen im Web entwickelt, der auf einem *Grundbaustein* und einem *Grundprinzip*, nach dem diese miteinander verbunden werden können, fusst:³ Den Grundbaustein stellen die Uniform

¹ Vgl. Berners-Lee et al.: Semantic Web.

² Ebd.

³ Bei den folgenden Ausführungen stütze ich mich im Wesentlichen auf Christian Stein: Linked Open Data – Wie das Web zur Semantik kam, in; Bibliothek, Forschung und Praxis 38 (2014), Heft 3, S. 1–9. https://www.interdisciplinary-laboratory.hu-berlin.de/de/content/linked-open-data-wie-das-web-zur-semantik-kam/?content_type=mitglied; Johannes Vetter: Semantik Web im Überblick. https://www.researchgate.net/publication/267845680_Semantic_Web_im_Ueberblick.

Resource Identifier (URI) dar. Mit ihnen werden alle basalen Daten oder „Ressourcen“ nach einer vorgegebenen Syntax eindeutig bezeichnet.⁴² Im Prinzip kann mehr oder weniger alles als Ressource fungieren und damit per URI referenziert werden: reale Lebewesen, fiktive Charaktere, abstrakte Ideen, Theorien und Konzepte, Gegenstände, Handlungen usw.

Das komplementäre Grundprinzip zur URI stellen die Triple dar. Damit ist vorerst nur gemeint, dass Aussagen, im Semantic-Web-Kontext als Statements bezeichnet, immer aus drei Bestandteilen bestehen müssen. Auch die Struktur, nach der diese Bestandteile zu einem Statement zusammengesetzt werden, ist vorgegeben. Sie wird wahlweise als Knoten-Kanten-, Ecken-Bögen- oder Subjekt-Prädikat-Objekt-Struktur beschrieben. Unabhängig von der Bezeichnung besteht das Grundprinzip der Struktur immer darin, dass ein Statement mit einer als Subjekt (respektive einem ersten Knoten) fungierenden Ressource beginnt und einem Objekt (respektive einem zweiten Knoten) endet. Die Verbindung zwischen beiden wird durch eine weitere Ressource, das Prädikat, respektive die Kante, hergestellt. Während es sich bei den ersten beiden Ressourcen, also Subjekt und Prädikat, um URIs handeln muss, kann das Objekt eine URI oder schlicht ein Literal sein. Entscheidend ist nun, dass alle Subjekte und Objekte wiederum als Subjekte oder Objekte eines neuen Statements fungieren können. Diese Graphstruktur, die einzelne Knotenpunkte in einer relativen flexiblen Weise zueinander in Beziehung setzt, schafft die Voraussetzung dafür, dass Ressourcen zu einem riesigen und komplexen Netzwerk aus Daten verbunden werden können, deren Semantik in einer maschinenlesbaren Form expliziert worden sind.

Wie bereits angeklungen ist, stellt ein aus URIs zusammengesetztes Triple nur das einfachste Strukturelement des anvisierten Datennetzes dar. Um die semantische Mächtigkeit noch weiter zu erhöhen, war von Beginn an auch die Verwendung von Ontologien angedacht. Der Begriff der Ontologie stammt ursprünglich aus einem philosophischen Kontext, in dem er die ‚Lehre vom Seienden‘ bezeichnet. Seit den Siebzigerjahren sind Ontologien erstmals auch in der Informatik, verstärkt seit den Neunzigerjahren insbesondere im Bereich der Entwicklung von Künstlicher Intelligenz, gebräuchlich, wenngleich mit der Übernahme der Begrifflichkeit auch eine Verschiebung der Bedeutung einherging: Zwar existiert in der Informatik und den an sie angrenzenden oder an sie anschließenden Disziplinen wie den Digital Humanities eine Vielzahl von teils divergierenden Auffassungen, Konsens dürfte aber darüber bestehen, dass unter Ontologien Konzepte zu verste-

⁴² <https://tools.ietf.org/html/rfc3986> [13.02.2019]

hen sind, die in Form formaler Begriffssysteme Wissen über einen spezifischen Gegenstandsbereich modellieren.⁴³ Sie sind, um auf die wohl bekannteste Definition zurückzugreifen, explizite Spezifikationen von Konzeptualisierungen, also abstrakten und vereinfachten Repräsentationen eines spezifischen Weltausschnitts.⁴⁴ Dabei muss darauf hingewiesen werden, dass Konzeptualisierungen immer auf geteiltem Wissen, also auf dem innerhalb einer Gruppe über den jeweiligen Wissensgegenstand erzielten Konsens, beruhen.

Da Ontologien vor allem die Kommunikation zwischen Mensch und Maschine erleichtern sollen,⁴⁵ besteht ihre zentrale Fragestellung mithin darin, „wie menschliches Wissen regelbasiert beschrieben und repräsentiert werden kann, um es einer maschinenverarbeitbaren Speicherung, Auswertung und Nachnutzung zugänglich zu machen.“⁴⁶ Dafür werden in Ontologien mithilfe verschiedener Komponenten formale Wissensmodelle definiert. Zu den Komponenten gehören⁴⁷:

- 1.) Klassen (classes). Als Klasse kann die Zusammenfassung von Dingen, die bestimmte Eigenschaften teilen, unter einem abstrakten Begriff (concept) verstanden werden. [Klassen können untereinander hierarchisch angeordnet werden];
- 2.) Attribute und Eigenschaften bzw. Merkmale (attributes, properties), die die Charakteristika der jeweiligen Klassen näher spezifizieren;
- 3.) Relationen bzw. Beziehungen (relations), die das Verhältnis zweier Klassen oder Instanzen zueinander charakterisieren;

⁴³ Definition nach Malte Rehbein: Ontologien, in: Fotis Jannidis, Hubertus Kohle, Malte Rehbein (Hgg.), Digital Humanities. Eine Einführung, Stuttgart 2017, S. 162–176, hier S. 162.

⁴⁴ „A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. [...] An ontology is an explicit specification of a conceptualization.“; Thomas R. Gruber: A Translation Approach to Portable Ontology Specifications, in: Knowledge Acquisition 5 (1993), Issue 2, S. 199-220, hier S. 199.

⁴⁵ Das Verhältnis Mensch/Maschine steht oft im Mittelpunkt des Interesses, wenn im informationswissenschaftlichen Kontext die Rede von Ontologien ist. Allerdings können Ontologien gleichfalls die Kommunikation innerhalb der jeweiligen ‚Sphären‘, also in der Mensch-Mensch- ebenso wie in der Maschine-Maschine-Kommunikation, erleichtern.

⁴⁶ Stefan Münnich: Ontologien als semantische Zündstufe für die digitale Musikwissenschaft? Eine Bestandsaufnahme, in: Bibliothek, Forschung und Praxis 42 (2018), Heft 2, S. 184–193, hier S. 185. DOI: 10.18452/18946.

⁴⁷ Die folgende Zusammenfassung der Ontologie-Komponenten nach Rehbein (2017), Ontologien.

4.) Bedingungen bzw. Restriktionen (constraints), die die Verwendung von Attributen und Eigenschaften reglementieren und so die logische Konsistenz der Ontologie sichern sollen;

5.) Grundannahmen (axioms), in denen Aussagen festgehalten werden, die für die jeweilige Ontologie immer korrekt sind, sich aber nicht aus den Begriffen selbst ergeben

6.) Instanzen oder Individuen, also die Objekte, bei denen es sich um die konkreten Ausprägungen der Konzepte handelt und die eine Ontologie erst ‚bevölkern‘.

Um die Idee des Semantic Web umsetzen zu können, wurden bereits seit den späten Neunzigerjahren, maßgeblich unter der Federführung des W3-Konsortiums, technische Standards für die Formulierung von Statements und Ontologien entwickelt. Ein solcher Standard ist z.B. das Resource Description Framework (RDF).⁴⁸ Ursprünglich als systemunabhängiger Metadatenstandard entwickelt, stellt RDF ein Vokabular zur Beschreibung von Ressourcen und ihren Beziehungen untereinander zur Verfügung. Mithilfe dieses Vokabulars lässt sich unter anderem ausdrücken, dass es sich bei Ressourcen um bestimmte Datentypen (z.B. `rdf:XMLLiteral` für Literale, Properties (`rdf:Property`) oder Instanzen einer Klasse (`rdf:type`) handelt. Darüber hinaus können die in RDF formulierten Statements auf verschiedene Arten - bisweilen auch als ‚Dialekte‘ bezeichnet - notiert, respektive repräsentiert werden; am gebräuchlichsten sind hier die Serialisierungen RDF/XML⁴⁹, Notation3 (N3)⁵⁰, Turtle⁵¹ und JSON-LD.⁵²

Im Fokus von RDF stehen im Wesentlichen ‚nur‘ Aussagen über Ressourcen, genauer gesagt über Instanzen und ihre Beziehungen zueinander. Um jedoch zu beschreiben, dass diese Ressourcen als Bestandteile einer bestimmten Gesamtheit aufzufassen sind, stellt RDF mit Ausdrucksmitteln wie `rdf:type`, `rdf:XMLLiteral`, `rdf:List`, `rdf:Statement` und anderen nur wenige Möglichkeiten bereit. Aus diesem Grund veröffentlichte das W3-Konsortium 1998 das Resource Description Framework Schema (wahlweise RDF(S), RDFS oder auch RDF-S).⁵³ Bei RDFS handelt es sich um eine Ontologiesprache. Ihr Ziel besteht also in der Bereitstellung möglichst universell verwendbar-

⁴⁸ <https://www.w3.org/RDF/> [12.02.2019]

⁴⁹ <https://www.w3.org/TR/rdf-syntax-grammar/> [12.02.2019]

⁵⁰ <https://www.w3.org/TeamSubmission/n3/> [12.02.2019]

⁵¹ <https://www.w3.org/TR/turtle/> [12.02.2019]

⁵² <https://json-ld.org> [12.02.2019]

⁵³ <https://www.w3.org/TR/rdf-schema/> [12.02.2019]

er Begriffe, auf deren Grundlage Ressourcen typisiert und (einfache) Ontologien erstellt werden können. So lassen sich in RDFS nicht nur mithilfe des Bezeichners `rdfs:Class` Klassen definieren, sondern auch durch `rdfs:subClassOf` bzw. – auf der Ebene der Relationen – `rdf:subPropertyOf` zu einfachen hierarchischen Komplexen anordnen. Bei diesen Hierarchien handelt es sich insofern um Taxonomien, als alle Unterklassen die Eigenschaften ihrer jeweiligen Oberklasse „erben“. Darüber hinaus bietet RDFS die Möglichkeit, die terminologischen Zusammenhänge zwischen Klassen und Properties genauer festzulegen, und zwar durch die Beschränkung des Definitionsbereichs einer Property mittels `rdfs:domain` auf mögliche Subjektklassen sowie des Wertebereichs einer Property mittels `rdfs:range` auf mögliche Objektklassen.

Damit ist RDFS, anders seine 'Grundlage' RDF, nicht nur in der Lage, zur formalen Beschreibung spezifischer Wissensbereiche Vokabulare zu nutzen, sondern diese auch selbst zu generieren. Wie bereits angedeutet wurde, ist die Komplexität der auf dieser Grundlage gebildeten Ontologien begrenzt. Um die semantische Expressivität noch weiter zu erhöhen, wurde, ebenfalls unter Federführung des W3-Konsortiums, die auf der Beschreibungssprache DAM+OIL basierende Web Ontologie Language (OWL) entwickelt, 2004 eingeführt und zu einer aus drei Ausprägungsstufen oder Spezies (OWL Lite, OWL DL, OWL Full) bestehenden Sprachenfamilie erweitert.⁵⁴ Im Gegensatz zu RDFS erlaubt OWL etwa Aussagen über Kardinalitäten und ermöglicht auf einer Beschreibungslogik beruhende maschinelle Inferenzen, also den Zugriff auf implizites Wissen mittels Schlussfolgen.

Darüber hinaus waren für die Entwicklung des Semantic Web weitere Standards von Bedeutung, von denen an dieser Stelle aber nur die beiden für die folgenden Ausführungen wichtigsten kurz erwähnt werden sollen. Dies ist zum ersten das Simple Knowledge Organization System (SKOS).⁵⁵ Mithilfe von SKOS lassen sich Taxonomien, Thesauri und andere Systeme zur Organisation von Wissen formalisieren und aufeinander abbilden oder – im Sinne des Semantic Web – harmonisieren. Zum zweiten ist dies SPARQL Protocol And RDF Query Language (SPARQL)⁵⁶, eine Abfragesprache auf der Basis von Graphen, mit der sich als RDF formulierte Triple über verschiedene Datenquellen hinweg abfragen lassen. Diese und andere, an dieser Stelle aus Platzgründen nicht vorgestellte Standards

⁵⁴ https://www.w3.org/standards/techs/owl#w3c_all [12.02.2019]

⁵⁵ <https://www.w3.org/TR/skos-reference/> [12.02.2019]

⁵⁶ <https://www.w3.org/TR/rdf-sparql-query/> [12.02.2019]

bilden zusammen den Semantic Web Stack, ein System aufeinander aufbauender Bausteine, das die Grundlage des Semantic Web darstellt.

Die Gebiete, in denen nun Semantic-Web-Technologien in der einen oder anderen Form in einem fortgeschrittenen Stadium erprobt oder bereits erfolgreich angewandt werden, sind vielfältig. So erfreuen sich beispielsweise Ontologien in den Lebenswissenschaften oder auch auf dem kommerziellen Sektor, insbesondere im betrieblichen Wissensmanagement, großer Beliebtheit. Das bekannteste Beispiel für den Einsatz graphbasierter Technologien dürfte sicherlich Googles Knowledge Graph sein.

Darüber hinaus haben Prinzipien und Technologien, die auf den Ansätzen der Semantic-Web-Idee beruhen, auch in verschiedenen Ausprägungen in Museen, Bibliotheken und ähnlichen Institutionen Fuß gefasst. Da diese Kulturerbeeinrichtungen thematisch und teilweise auch institutionell mit dem (digitalen) Editionsweisen verbunden sind und – wie noch zu zeigen sein wird – im Hinblick auf Semantic-Web-Technologien ein infrastrukturelles Umfeld entwickelt haben, das für digitale Editionsprojekte einen natürlichen Anknüpfungspunkt bietet, soll im Folgenden näher auf die Entwicklungen in diesem Bereich eingegangen werden.

Mit Blick auf das Museumswesen dürfte sicherlich die Entwicklung des CIDOC Conceptual Reference Model (CRM)⁵⁷ als eine Erfolgsgeschichte gelten: Das CIDOC CRM wird seit 1996 unter Federführung des Internationalen Museumsrates ICOM als Application Ontology für die Dokumentation von musealen Artefakten entwickelt und wurde 2006 sogar als ISO-Standard für die Datenmodellierung im Kulturerbebereich fixiert.⁵⁸ Allerdings bleibt ihre Anwendung nicht auf die reine Museumsdokumentation beschränkt. Wie die Entwickler selber betonen (und auch die Namensgebung als „Reference Modell“ signalisiert), soll das CIDOC CRM als transdisziplinäre „lingua franca“ zur Vereinheitlichung von Wissen zu Kulturerbeobjekten und -strukturen dienen.

Auf dieser Basis eines harmonisierten maschinenprozessierbaren Kulturwissens soll CIDOC CRM ebenso den Austausch von relevanten Informationen zwischen Museen, Bibliotheken, Archiven und vergleichbaren Institutionen ermöglichen wie zwischen verschiedenen Forschungsprojekten, solange sich diese im weitesten Sinne dem Gebiet der Kulturerbeerschließung zurechnen lassen. In der Tat fungiert das CIDOC CRM dank sein-

⁵⁷ Für die aktuelle Version vgl. http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf [17.02.2019]

⁵⁸ <https://www.iso.org/standard/57832.html> [17.02.2019]

er ausgereiften und komplexen Struktur – die Zahl der Klassen beläuft sich auf 90, die der Properties auf knapp 150 – als eine Art „Referenzontologie“⁵⁹, aus der sich wiederum nach Bedarf (Zuschnitt des jeweiligen Projekts, Forschungsdisziplin, Art der dokumentierten Objekte, Erschließungstiefe usw.) spezifische Anwendungs- oder Domänenontologien ableiten lassen.⁶⁰

Auch im Bibliothekswesen haben sich Semantic-Web-Technologien in verschiedenen Ausprägungen etabliert. Am weitesten ist diese Entwicklung wohl im Bereich der Anwendung von Linked Open Data für Katalogdaten gediehen.⁶¹ Angesichts der technischen Entwicklungen seit den Neunzigerjahren wurde hier schon vergleichsweise früh von verschiedenen Seiten Kritik an den gängigen bibliotheksarischen Standards geübt⁶², für die in der Folgezeit alternative Semantic-Web-fähige Formate entwickelt wurden. Der wohl bedeutendste Schritt in dieser Richtung war sicherlich die Formulierung der Functional Requirements for Bibliographic Records (FRBR) sowie die Einführung des darauf basierenden Regelwerks Resource Description and Access (RDA): FRBR wurde bereits seit 1998 auf Basis einer eigenen Ontologie unter Federführung der IFLA entwickelt und stellt heute den wohl wichtigsten theoretischen Bezugsrahmen für die Formulierung bibliothekarischer Regelwerke dar, von denen wiederum RDA das mit Abstand einflussreichste ist. Nach

⁵⁹ Georg Hohmann: Die Anwendung von Ontologien zur Wissensrepräsentation und -kommunikation im Bereich des kulturellen Erbes, in: Schomburg, Silke; Leggewie, Claus; Lobin, Henning; Puschmann, Cornelius (Hgg.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland, Köln 2011, S. 33–39, hier S. 36.

⁶⁰ Vgl. z.B. Laura Albers, Peggy Große, Sarah Wagner: Semantische Datenmodellierung mit CIDOC CRM – Drei Fallbeispiele, in: Maximilian Eibl, Martin Gaedke (Hgg.), Informatik 2017. Tagung 2017 in Chemnitz, Band 2 [= Jahrestagung der Gesellschaft für Informatik 47], Bonn 2017, S. 1209–1220. DOI: <https://doi.org/10.11588/artdok.00005745>. Der Beitrag illustriert exemplarisch, wie sich auf CIDOC CRM fußend für Forschungsdaten aus drei denkbar unterschiedlichen Themengebieten (die – v.a. begriffliche – Repräsentation abstrakter Friedenskonzepte, frühneuzeitliche Kunst- und Wunderkammern sowie romanische Wandmalereien) Datenmodelle ‚zuschneiden‘ lassen.

⁶¹ Für einen Überblick über die Entwicklung im deutschsprachigen Raum vgl. z.B. Adrian Pohl, Patrick Danowski: Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick, in: dies. (Hgg.), (Open) Linked Data in Bibliotheken, Berlin, Boston 2013, S. 1–44. DOI: 10.1515/9783110278736.1. Ähnlich gelagert, allerdings mit einem Akzent auf dem angelsächsischen Raum: Yongming Wang, Sharon Q. Yang: Linked Data Technologies and What Libraries Have Accomplished So Far, in: International Journal of Librarianship, 3 (2018), Issue 1, S. 3–20. DOI: <https://doi.org/10.23974/ijol.2018.vol3.1.62>.

⁶² Nachgerade berühmt in diesem Zusammenhang ist beispielsweise Roy Tennants programmatisch betitelter Beitrag „MARC must die“, in dem er die auf MARC und AACR(2) basierende Erfassung von Titeldaten als „paper catalog card in computer form“ bezeichnet, denen einerseits die notwendige Granularität fehle und deren hochkompliziertes Regelwerk nur von einer überschaubaren Gruppe ausgewiesener bibliothekarischer Fachleute gehandhabt werden könne; vgl. Roy Tennant: MARC must die, in: Library Journal, 127, Nr. 17, S. 26–27.

seiner ersten Veröffentlichung im Jahre 2010 löste es sukzessive sowohl die in Deutschland und Österreich gebräuchlichen Regeln für die alphabetische Katalogisierung (RAK) als auch ihr im angelsächsischen Raum gebräuchliches Pendant, die Anglo-American Cataloguing Rules (AACR) ab. Darüber hinaus entwickelte die Library of Congress mit dem Bibliographic Framework (BIBFRAME)⁶³ ein in RDF verfasstes Datenmodell, das dezidiert zur „Implementierung von RDA und anderen Regelwerken mit den Mitteln von Linked Data/Linked Open Data“⁶⁴ konzipiert wurde und die bisherigen Standard-Formate, insbesondere MACHine-Readable Cataloging (MARC), aber auch das Metadata Object Description Schema (MODS) ersetzen soll.

Die Anwendungsszenarien von Semantic-Web-Technologien beschränken sich aber nicht allein auf die Konversion von Katalog-Altdaten. Schon vor der Formulierung der Grundprinzipien von Linked Open Data durch Tim Berners-Lee erregten beispielsweise die Potentiale eines auf Ontologien beruhenden Information Retrieval und der Datenharmonisierung mittels semantischer Datenmodelle die Aufmerksamkeit von Teilen der bibliothekarischen Community.⁶⁵ Ein Katalysator für das wachsende Interesse waren die seinerzeit angedachten oder bereits in der Entwicklung befindlichen digitalen und virtuellen Bibliotheken. Diesem Bereich ist denn auch eines der wohl ambitioniertesten Projekte in diesem Zusammenhang zuzuordnen: die Europeana.⁶⁶ Die Europeana wurde seit 2005 als zentrales digitales Nachweisportal für das europäische Kulturerbe entwickelt. Wichtig ist in diesem Zusammenhang, dass es sich bei ihr um eine virtuelle Bibliothek handelt. Sie bildet also nicht den Bestand einer realen Bibliothek digital ab, sondern führt „nur“ die Metadaten zu den Objekten verschiedener Institutionen zusammen, ohne diese physisch zu verwalten. Entsprechend stand die der Europeana zugrundeliegende Informationsarchitektur, das Europeana Data Modell (EDM)⁶⁷, vor der Aufgabe, die teils sehr unter-

⁶³ Nach der 2012/13 präsentierten ersten Version wurde 2016 mit BIBFRAME 2.0 die aktuelle Version des Rahmenwerks vorgelegt.

⁶⁴ Julia Hauser et al.: Interoperable Metadaten im Giant Global Graph, in: Dialog mit Bibliotheken 2014, Heft 2, S. 41–46, hier S. 45.

⁶⁵ Vgl. als Übersicht und frühes Beispiel zugleich Sabine Boltzendorf: Ontologien in digitalen Bibliotheken unter dem Schwerpunkt Inhaltserschließung und Recherche [= Berliner Handreichungen zur Bibliothekswissenschaft 111]. Berlin 2013. DOI: 10.18452/18353.

⁶⁶ <https://www.europeana.eu/portal/de> [19.02.2019]

⁶⁷ <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>; für eine ausführliche Darstellung vgl. Martin Doerr, Stefan Gradmann, Steffen Hennicke et al.: The Europeana Data Model (EDM). IFLA 2010, World Library and Information Congress: 76th IFLA General Conference and Assembly, 2010, Gothenburg. URL: <<https://www.ifla.org/past-wlic/2010/149-doerr-en.pdf>>.

schiedlichen „Objektrepräsentationen unterschiedlicher Inhaltsanbieter zu einem Netzwerk von ‚Resources‘ zu verbinden“⁶⁸ – ein alles andere als triviales Unterfangen angesichts von derzeit gut 50 Millionen in der Europeana verzeichneten Objekten, deren Metadaten von hunderten europäischen Institutionen aus den verschiedensten Wissensdomänen und Kulturerbesparten geliefert werden und die neben Büchern u.a. auch Handschriften, Bilder (Fotos, Gemälde, Zeichnungen usw.) sowie Audio- und Videomaterialien umfassen. Um die Bestände dennoch möglichst konsistent zusammenzuführen, wurde das EDM – nach Möglichkeit unter Rückgriff auf existierende Standards und Formate wie SKOS, Dublin Core und natürlich RDF(S) – als Top-level-Ontologie entwickelt, die „eine Reihe generischer Attribute und Klassen an[bietet], die als Anker für die spezifischeren Beschreibungen der Datenprovider dienen. Damit werden die ursprünglichen Datenrepräsentationen weitestgehend erhalten und gleichzeitig Interoperabilität gewährleistet.“⁶⁹ Auf einem ähnlichen Ansatz beruht auch die Deutsche Digitale Bibliothek (DDB)⁷⁰, die die Metadaten zu momentan über 24 Millionen Objekten verschiedenster deutscher Kulturinstitutionen aggregiert. Dies geschieht dadurch, dass die aus den Objektdaten der Datenlieferanten gewonnenen Informationen zu RDF-Tripeln vereinheitlicht und auf eine Ontologie gemappt werden, die auf dem CIDOC CRM beruht.

Wie in den obigen Ausführungen bisweilen schon angeklungen ist, liegt ein entscheidender Grund für die Attraktivität von Semantic-Web-Technologien sicherlich in der Hoffnung, dass sich mit ihrer Hilfe Barrieren zwischen den einzelnen Kulturerbesparten überwinden und in einem semantisch harmonisierten Wissensraum zusammenführen lassen. Und in der Tat können Projekte wie die Europeana und die Deutsche Digitale Bibliothek bis zu einem gewissen Grad diese Hoffnungen auch einlösen, da sie auf Grundlage ihrer Datenmodelle Bestände zusammenführen, die in vielerlei Hinsicht – Medienformat, Metadatenformat, aufbewahrende Kulturerbesparte usw. – äußerst heterogen sind. Darüber hinaus existieren auch einige Arbeitsgruppen und andere Initiativen, die sich eine weitere Harmonisierung der unterschiedlichen Standards zum Ziel gesetzt oder bereits umgesetzt hat. Die Anfänge eines der wichtigsten Projekte in dieser Hinsicht, die Functional Require-

⁶⁸ Esther Chen: Linked Open VD 17 – von METS/MODS zum Europeana Data Model. Überlegungen zum technischen Migrationspfad und zum funktionalen Mehrwert semantischer Nutzungsszenarien [= Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft 327]. Berlin 2012, S. 17. DOI: 10.18452/2069.

⁶⁹ Steffen Hennicke: Linked Data und semantische Suchfunktionalität in Europeana, in: Mitteilungen der VÖB 66 (2013), Heft 1, S. 20–34, hier S. 20.

⁷⁰ <https://www.deutsche-digitale-bibliothek.de> [19.02.2019]

ments for Bibliographic Records - object oriented (FRBRoo)⁷¹, die als Mediations-, respektive Austauschformat zwischen FRBR und dem CIDOC CRM entwickelt wurde, datieren bereits auf das Jahr 2006. Allerdings ist auch trotz solcher Initiativen die antizipierte Interoperabilität alles andere als ein Selbstläufer. Dies liegt sicherlich unter anderem an spartenspezifischen Traditionen und Kulturen, wie ein Blick auf RDA und FRBR deutlich macht: Ursprünglich wurde RDA von bibliothekarischer Seite dezidiert mit dem Anspruch entwickelt, für möglichst viele, wenn nicht alle Kulturerbe-Institutionen verwendbar zu sein. Allerdings wurde RDA ebenso wie das ihm zugrundeliegende FRBR sowohl von Museums- als auch von Archivseite als wenig praktikabel bezeichnet. Dies richtete sich vor allem gegen das Herzstück der FRBR, die ‚Auffächerung‘ eines Objekts in die Konzepte ‚Werk‘, ‚Manifestation‘, ‚Expression‘, ‚Exemplar‘, die wiederholt als ‚zu bibliothekarisch‘ gedacht und für unikale Objekte – und mit diesen haben es Archive wie Museen natürlich vornehmlich zu tun – ungeeignet kritisiert.⁷²

Dessen ungeachtet sind die Fortschritte in der Vernetzung der Bestände zwischen den einzelnen Kulturerbesparten und über sie hinaus nicht zu unterschätzen. Das liegt auch daran, dass sich neben den expressiven, schwergewichtigen Ontologien wie CIDOC CRM auch ‚niedrigschwelligere‘ Ansätze wie die Verknüpfungen auf der Instanzebene als effektiver semantischer Transmissionsriemen herausgestellt haben. Die Grundlage hierfür bildete das enorme Wachstum an Semantic-Web-fähigen Datensammlungen: Neben den umfassenden und als eine Art ‚Grundbaustein‘ der LOD-Cloud fungierenden Datensets wie DBpedia und GeoNames wurden im Laufe der letzten Jahre immer mehr für den Kulturerbebereich relevante Sammlungen in einer RDF-Version veröffentlicht und somit der Community zugänglich gemacht. Die Gemeinsame Normdatei hat sich, nicht zuletzt aufgrund ihrer Datenqualität und ihres Umfangs von momentan rund 153 Millionen RDF-Tripeln⁷³, zum bedeutendsten Baustein in diesem Zusammenhang entwickelt. Die GND

⁷¹ <https://www.ifla.org/node/10171> [19.02.2019]

⁷² Je nach Kulturerbesparte und damit zusammenhängender Perspektive auf die verwalteten Objekte werden neben diesem Hauptkritikpunkt auch fehlende Möglichkeiten zur Angabe von Provenienz und Materialität (betrifft v.a. Kritik von Museumsseite) oder zur Einbindung in die Bestandstektonik (Archivseite) moniert; vgl. u.a. Irmgard Christa Becker: Archivische Erschließung und RDA. Stellungnahme der Archivschule Marburg zur Nutzung der RDA (Resource Description and Access) in Archiven. http://www.dnb.de/SharedDocs/Downloads/DE/DNB/standardisierung/rdaKultur2013Becker.pdf?__blob=publicationFile; Christian Aliverti, Claudia Fabian, Angela Kailus: RDA und Kultureinrichtungen, in: Zeitschrift für Bibliothekswesen und Bibliographie, 62 2015, Heft 6, S. 329-338. <http://dx.doi.org/10.3196/186429501562643>.

⁷³ <https://lod-cloud.net/dataset/dnb-gemeinsame-normdatei> [01.03.2019]

entstand 2012 unter der Federführung der Deutschen Nationalbibliothek durch die Zusammenlegung der Normdateien für Personen (PND), Körperschaften (GKD) und Schlagworte (SWD). Seit 2014 erfolgt die Erfassung auf der Grundlage von RDA. Schon vier Jahre zuvor wurde damit begonnen, die Daten nicht nur ‚klassisch‘ in einer MARC-Kodierung, sondern auch als Linked Open Data in einem RDF-Format anzubieten. Grundlage dafür ist eine eigene Ontologie⁷⁴, die neben eigenen Elementen nach Möglichkeit auf die Verwendung von oder das Alignment mit etablierten Vokabularen wie FOAF setzt, um größtmögliche Kompatibilität zu garantieren. Aufgrund seiner Quantität und seiner Qualität – der Großteil der GND-Daten wird intellektuell erfasst und gepflegt – stellen die GND zusammen mit bereits genannten Datensets wie GeoNames den bedeutendsten semantischen Knotenpunkt des deutschsprachigen Raums zu Personen, Körperschaften, Veranstaltungen, Orten, Werken und Schlagwörtern dar. Wie bereits angedeutet, existiert darüber hinaus noch eine Vielzahl weiterer Projekte, die – oft auf älteren relationalen Datenbanken aufbauend – für den Kulturerbebereich relevante Datensammlungen als Linked Open Data zur Verfügung stellen. Diese Entwicklung in ihrer Gesamtheit darzustellen, würde den Rahmen dieser Arbeit sprengen. Pars pro toto seien an dieser Stelle mit dem Professorenkatalog der Universität Leipzig⁷⁵, der Deutschen Biografie⁷⁶ und dem Projekt Census⁷⁷ drei in ihrer Unterschiedlichkeit die Breite der verfügbaren Daten illustrierende Beispiele zumindest kurz erwähnt.

Zusammenfassend bleibt festzuhalten, dass sich trotz diverser Probleme und Kritikpunkte – entsprechende Diskussionen konnten hier nur im Zusammenhang mit RDA und FR-BRoo kurz angedeutet werden – in den letzten Jahren im Kulturerbebereich ein umfassende Semantic-Web-fähige Struktur aus verschiedenen Instrumenten (Modelle, Vok-

⁷⁴ <https://d-nb.info/standards/elementset/gnd> [19.02.2019]

⁷⁵ <https://research.uni-leipzig.de/catalogus-professorum-lipsiensium/>; zur technischen Realisierung, unter anderem zur eigens entwickelten Ontologie *Catalogus Professorum Model*, vgl. Christian Augustin, Ulf Morgenstern, Thomas Riechert: Der Leipziger Professorenkatalog. Ein Anwendungsbeispiel für kollaboratives Strukturieren von Daten und zeitnahes Publizieren von Ergebnissen basierend auf Technologien des Semantic Web und einer agilen Methode des Wissensmanagements, in: Schomburg, Silke; Leggewie, Claus; Lobin, Henning; Puschmann, Cornelius (Hgg.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland, Köln 2011, S. 41–44.

⁷⁶ <https://www.deutsche-biographie.de/> [21.02.2019]

⁷⁷ Zu Generierung von und Arbeit mit Linked Open Data im Rahmen des Projekts vgl. <http://www.census.de/census/nachrichten/kulturerbe-als-linked-open-data-pilotprojekt-fuer-den-census>; zum Projekt selber vgl. <http://www.census.de/census> [21.02.2019]

abulare, Regelwerke usw.) und darauf beruhenden Datensammlungen entwickelt hat. Das so entstandene Ensemble ist domänenübergreifend und umfassend, bewahrt aber trotz seiner überbordenden quantitativen Ausmaße dank der Grundprinzipien des Semantic Web eine gewisse Konsistenz und Integrität. Für unseren Zusammenhang ist entscheidend, dass dieses Korpus ein Semantic-Web-fähiges Bezugssystem zur Verfügung stellt, innerhalb dessen sich alle möglichen Formen von historisch-kulturellen Informationen synergetisch verbinden und austauschen lassen. Damit stellt es einen ideale Anknüpfungspunkt für Editionsprojekte auf Semantic-Web-Basis dar. Ob und in welcher Form diese Möglichkeit der Anknüpfung genutzt wird, soll Gegenstand der folgenden Kapitel sein.

3. Semantic-Web-Technologien & digitale Editionen – Theoretische Ansätze

Als mehr oder weniger unangefochtener Standard für das digitale Edieren haben sich die Richtlinien der Text Encoding Initiative (TEI) etabliert.⁷⁸ Die Grundidee der TEI besteht darin, ein über die Grenzen verschiedener Fach-Disziplinen anwendbares Ensemble aus digitalen Markup-Elementen sowie Regeln für deren Anwendung bereitzustellen. Seit ihren Anfängen im Jahre 1987 hat sich die TEI, die von einem aus verschiedenen internationalen Forschungseinrichtungen zusammengesetzten Konsortium getragen und beständig weiterentwickelt wird, zu einem Normenwerk entwickelt, das neben jährlichen Konferenzen auch eine eigene Zeitschrift, das mittlerweile schon in der 13. Ausgabe erschienene *Journal of the Text Encoding Initiative*,⁷⁹ unterhält. Kernstück sind die Guidelines der TEI, die in ihrer aktuellen Form, dem Proposal 5, aus rund 500 Elementen sowie dazugehörigen Erläuterungen besteht.⁸⁰

Auf dieser Grundlage lassen sich Metadaten ebenso wie Daten auszeichnen. Dafür wird generell zwischen dem Header und dem eigentlichen Quelltext unterschieden. Ersterer kann bibliographische Angaben wie Titel (ausgezeichnet mit dem Tag `title`), Verfasser (`<author>`), Erscheinungsort und -jahr (`<pubPlace>` bzw. `<date>`), aber auch eine allgemeine Beschreibung der Vorlage (`<sourceDesc>`) enthalten. Für die Kodierung des Quelltextes halten die TEI-Guidelines neben allgemein gültigen *core elements*, etwa `<p>` für Absätze, diverse spezifische Module bereit, etwa für Wörterbücher oder für paläographische Anmerkungen. Darauf aufbauend lässt sich der Text im Hinblick auf eine ganze Reihe von Gesichtspunkten – etwa Textstruktur, Textvarianten und -genese, linguistische Phänomene, aber auch Entitäten wie Ort und Person usw. – auszeichnen.

Zwar ist die Entwicklung der TEI, die nicht nur in der digitalen Editorik, sondern auch als Grundlage für den Austausch von Handschriftenbeschreibungen, bei der Auszeichnung retrodigitalisierter historischer Buchbestände sowie bei virtuellen Forschungsumgebungen wie TextGrid und Ediarium zur Anwendung kommt, eine Erfolgsgeschichte. Allerdings wurden schon früh die Limitationen des SGML- bzw. XML-basierten Datenformats erkennbar.

Zum Ersten richtete sich die Kritik auf die dem Datenmodell zugrundeliegende *Ordered Hierarchy of Content Objects* (OHCO): Im Wesentlichen basiert die OHCO auf der Auffassung, dass ein Text als lineare Abfolge von Informationseinheiten oder *content objects*

⁷⁸ <http://www.tei-c.org> [02.04.2019]

⁷⁹ <https://journal.tei-c.org/journal/index> [02.04.2019]

⁸⁰ <http://www-old.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> [02.04.2019]

verstanden werden kann.⁸¹ Eine solches Textverständnis geht von dem Grundprinzip aus, dass unterschiedliche Informationseinheiten entweder aufeinander folgen, ohne sich zu überschneiden, oder aber ineinander hierarchisch geordnet werden können – ein Phänomen, das mit der Metapher des „Einnistens“ beschrieben wird. In der Tat lässt sich auf diese Weise beispielsweise ein Buch relativ problemlos als eine Abfolge von größeren abgeschlossenen Gliederungseinheiten (z.B. die Kapitel eines Romans oder die einzelnen Gedichte einer Lyriksammlung) strukturieren, in die jeweils kleinere Segmente (z.B. Absätze oder Verse) „eingenistet“ sind, die ihrerseits wieder ein Nest für noch kleinere Strukturelemente (z.B. Zeilen oder Strophen) bilden können. Zwar korreliert ein solches sequentielles Verständnis mit der Funktionsweise der TEI zugrundeliegenden Metasprache SGML bzw. XML, allerdings wird sie der Komplexität von Texten nicht gerecht. Dieses sogenannte Problem der Überlappung (bzw. *overlapping*) der Markups zu wird insbesondere dann relativ schnell deutlich, wenn die Auszeichnung des Textes in Hinblick auf verschiedene Phänomene vorgenommen wird, die sich durchaus nicht in eine singuläre und hierarchische Ordnung bringen lassen.

Angesichts zunehmender technischer Möglichkeiten hat sich das Problem der Überlappung in den letzten Jahren eher verschärft, denn während „dieser Umstand [i.e. die Überlappung] noch vor einiger Zeit als praktisch weniger relevant eingestuft wurde [...], kann er heute angesichts des Wunsches, immer mehr Annotationsebenen und Strukturierungsvarianten gleichzeitig digital abzubilden, nicht mehr ignoriert werden.“⁸²

Ein weiteres Problem stellt das sogenannte Over-tagging dar: Theoretisch bietet XML die Möglichkeit, den Text beliebig tief und umfassend zu erschließen und mit entsprechend vielen Markups zu versehen. Praktisch konfligiert diese Vorgehensweise mit dem Anspruch des TEI-Formats, nicht nur maschinen-, sondern auch menschenlesbar zu sein. Denn mit der Anzahl der Markups sinkt die Übersichtlichkeit des ursprünglichen Textes, und im extremsten Fall ist das Endergebnis ein von TEI-„Paratext“ regelrecht „zerfressener“ Originaltext. Wie beim überlappenden Markup ist auch beim Over-tagging anzunehmen, dass sich das Problem mit den zunehmenden Erwartungen, die an die Arbeit mit digital edierten Texten geknüpft werden, eher verstärken als verringern wird.

⁸¹ Vgl. Steven J. DeRose et al.: What is text, really? In: Journal of Computing in Higher Education 1,2 (1990), S. 3–26. <https://doi.org/10.1007/BF02941632>.

⁸² Thomas Efer: Graphdatenbanken für die textorientierten e-Humanities. Leipzig 2017, S. 36f. urn:nbn:de:bsz:15-qucosa-219122.

Weitere Probleme ergeben sich aus der Vielzahl der verfügbaren Tags. Denn die dadurch entstehenden vielfachen Kodierungsmöglichkeiten können dazu führen, dass ein und derselbe Inhalt mit verschiedenen Markups ausgezeichnet wird. So lässt sich ein im Text erwähnter Mensch sehr spezifisch mit `persName`, aber ebenso mit `name` oder generisch mit `rs` (referencing string) kodieren.⁸³ Schon relativ früh hat die TEI versucht, diesem Phänomen durch die Entwicklung eines reduzierten Tagsets (*TEI Lite*) zu begegnen, das als gemeinsamer Basis-Standard dienen soll, indem es „90% of the needs of 90% of the TEI user community“ erfüllt.⁸⁴ Allerdings können solche basalen Sets nicht die Tendenz der TEI-Benutzer zur sogenannten „Customization“, also der individuellen Auswahl von TEI-Tags, teilweise sogar angereichert mit projektspezifischem Markup, und deren Zusammenfassung in einem eigenen Schema⁸⁵, verhindern. Mit Blick auf die eingeschränkte Kompatibilität, die eine solche um sich greifende „Customization“ nach sich zieht, ist die Befürchtung mancher TEI-Anwender nicht ganz unberechtigt, dass „der TEI Standard [...], entgegen den Intentionen seiner Schöpfer, nicht zu einer Interoperabilität von Editionen, sondern vielmehr zu einer immer stärkeren Auffächerung des Markups geführt“⁸⁶ hat.

Darüber hinaus wird – um auf den letzten und für den Kontext dieser Arbeit vielleicht wichtigsten Kritikpunkt einzugehen – dem TEI-Format von verschiedenen Seiten attestiert, semantische Inhalte nicht oder – gemessen an heutigen Möglichkeiten – nur unbefriedigend explizieren zu können.⁸⁷ Der Malus der mangelhaften semantischen Expressivität hat seinen Ursprung unter anderem im Entstehungskontext der TEI: In der Traditionslinie der klassischen Editionsphilologie stehend, richteten die Initiatoren des Formats ihre Aufmerksamkeit in erster Linie auf formale Aspekte des Textes. In der Folge sind noch heute „die Richtlinien der Text Encoding Initiative (TEI) als De-facto-Standard digitalen Editierens z.B. mehr an komplexen Überlieferungsverhältnissen, kodikologischen und paläographischen Details oder an linguistischen Phänomenen interessiert als an der Erschließung von

⁸³ Beispiel nach Torsten Schaßan, Digitale Quellen – Datei- und Datenformate, in: Laura Busse et al. (Hgg.), Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften [= Historisches Forum 19], Berlin 2016, A.6. <http://www.clio-online.de/guides/arbeitsformen-und-techniken/digitale-quellen-dateiformate/2016>.

⁸⁴ <http://www.tei-c.org/guidelines/customization/lite/> [06.04.2019]

⁸⁵ Vgl. Schaßan, Digitale Quellen, A.6.

⁸⁶ Wettlaufer, Der nächste Schritt?, S. [1].

⁸⁷ Zur Kritik aus Sicht der TEI-„Community“ vgl. zusammenfassend Ciotti: Formal Ontologies, Linked Data, and TEI Semantics.

Propositionen.⁸⁸ Zwar lassen sich inhaltlich relevante Entitäten wie beispielsweise Personen, Orte oder Werktitel ebenfalls annotieren. Der semantische Gehalt solcher Auszeichnungen beschränkte sich aber lange Zeit darauf anzuzeigen, dass die markierte Zeichenfolge eine Person oder einen Ort usw. bezeichnet. Das bedeutet für die maschinelle Verarbeitung: Dergestalt ausgezeichnete Texte sind zwar theoretisch sowohl nach Obergruppen von Entitäten – also z.B. nach allen Städten, Flüssen usw. – als auch nach spezifischen Entitäten – also z.B. Personen mit dem Namen „Ludwig van Beethoven“ usw. – durchsuchbar, ihr konkreter semantischer Gehalt bleibt aber allein Sache menschlicher Interpretation. Letztere mag sich in natürlicher Sprache in Guidelines oder Lexemen wie dem Markup selbst niederschlagen. Für Maschinen sind diese Inhalte aber nicht oder nur eingeschränkt prozessierbar, denn im Wesentlichen bleiben sie darauf beschränkt, Texte formal zu verarbeiten. TEI-Markup ist also, anders gesagt, für Maschinen lesbar, aber nicht verständlich im Sinne des Semantic Web: „As far as an XML parser is concerned, `<title>Finnegans Wake</title>` and `<blob>Finnegans Wake</blob>` are both acceptable and well-formed markup instances.“⁸⁹

Lange waren damit TEI-kodierte Texte semantisch gesehen geschlossene Systeme, die zwar beispielsweise den Namen einer Entität markieren, diese Markierung aber mit keinen Informationen außerhalb des Textes verbinden konnten. Erst mit der Einführung des jüngsten, fünften Proposals der TEI wurde diese Barriere in mancher Hinsicht überwunden. So wurde eine Reihe von Annotationen zur Verbindung von Text- und externen Informationen eingeführt oder für diese Anforderungen erweitert. Trotz dieser und anderer Entwicklungen⁹⁰ bleibt als zentraler Kritikpunkt bestehen, dass die semantische Expressivität des TEI-Instrumentariums weit hinter den Möglichkeiten zurückbleibt, die die Entwicklung neuerer Technologien, insbesondere des Semantic Web, aufzeigen – ein Defizit, das umso schwerer wiegt, als die TEI in den digitalen Geisteswissenschaften im Allgemeinen und im digitalen Editionsweisen im Speziellen den Status eines De-facto-Standards genießen.

Im Folgenden sollen nun verschiedene theoretische und explorative Ansätze vorgestellt werden, die, an die Kritik der bisherigen Datenmodellierung anknüpfend, die Weiterentwicklung der Editionspraxis mithilfe von Semantic-Web-Technologien in ihren Mittelpunkt

⁸⁸ Vogeler, Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?, S. [2].

⁸⁹ Ciotti: Formal Ontologies, Linked Data, and TEI Semantics, S. 2.

⁹⁰ Vgl. beispielsweise die Ausführungen zu „TEI und Graphen“ in Jakob Šimek: Der Nutzen graph-basierter Datenmodelle für digitale Editionen. Unveröffentlichte Masterarbeit. Berlin 2018, S. 35-50.

gestellt haben. Nach Möglichkeit wird dabei auch ein Bezug zwischen den jeweiligen Konzepten und der im Eingangsteil dieser Arbeit vorgestellten Matrix hergestellt werden. Der Aspekt, der bei der Fachcommunity von Beginn an auf das größte Interesse stieß, war das Vernetzungspotenzial des Semantic Web. Der Mehrwert von eindeutigen und projektübergreifenden ‚Datenknotenpunkten‘ leuchtet unmittelbar ein: Schon früh wurde auf die Vorteile hingewiesen, die Normdatensätze wie die GND, die automatisch auch als URI referenziert sind, dadurch bieten, dass sie Personen und Orte eindeutig identifizierbar und so zweifelsfrei über die Grenzen der einzelnen Editionsprojekte hinweg ‚ansteuerbar‘ machen.⁹¹ In eine ähnliche Richtung, nur auf einer abstrakteren Ebene, weist auch die Möglichkeit, mithilfe von SKOS Terminologieprobleme wie die unterschiedliche Benennung ein und derselben Sache dadurch zu beheben, dass man die Benennungen zu einem gemeinsamen Konzept zusammenführt.⁹² Letzteres wäre dann, in eine URI überführt, eine ebenso projekt- und bestenfalls domänenübergreifend referenzierbare Ressource wie die URI eines GND-Personensatzes.

Mit der projektübergreifenden Referenzierbarkeit allein sind die Potenziale der Verwendung von URIs und anderen Semantic-Web-Standards für digitale Editionen allerdings noch nicht erschöpft. Vielmehr bildet sie lediglich die Grundlage für eine viel weitreichendere Operation, nämlich – wie es bei Iglesias et al. heißt – „die wissenschaftlich fundierte Suche nach und die disziplinen- wie spartenübergreifende Verbindung von Informationen [...]“. Die Vernetzung wiederum eröffnet neue Möglichkeiten, sich Informationen zu erschließen und stellt gleichzeitig ein Angebot an WissenschaftlerInnen (und im Falle von ‚Open‘ Data auch an viele andere Interessierte) dar, einerseits ihre eigenen Daten in Beziehung zu anderen Daten zu setzen, andererseits über den Horizont der eigenen Disziplin hinauszublicken und damit auch neue Perspektiven auf das eigene Fach zu entdecken. Die Verbindung von Informationen über semantische Beziehungen zeigt Zusammenhänge auf, die bei isolierter Betrachtung der

⁹¹ Vgl. als frühes Beispiel Peter Stadler: Normdateien in der Edition, in: *Editio* 26 (2012), Heft 1, S. 174–183. DOI: <https://doi.org/10.1515/editio-2012-0013>.

⁹² Ein Beispiel für ein fachspezifisches Terminologieproblem und dessen Lösung durch SKOS stellt Robert Kummer vor: In der kodikologischen Fach-Community ist es gängige Praxis, Materialarten in der jeweiligen Landessprache (also z.B. „Papier“ / „paper“ / „paipier“ oder „Pergament“ / „parchement“ usw.) anzugeben. Mit SKOS ließen sich „Papier“, „paper“ und „paipier“ durch `alt: Label` als unterschiedliche Benennungen desselben Konzepts kennzeichnen und mit einer URI zentral referenzieren; vgl. Kummer, *Semantic Technologies for Manuscript Descriptions*, S. 144f.

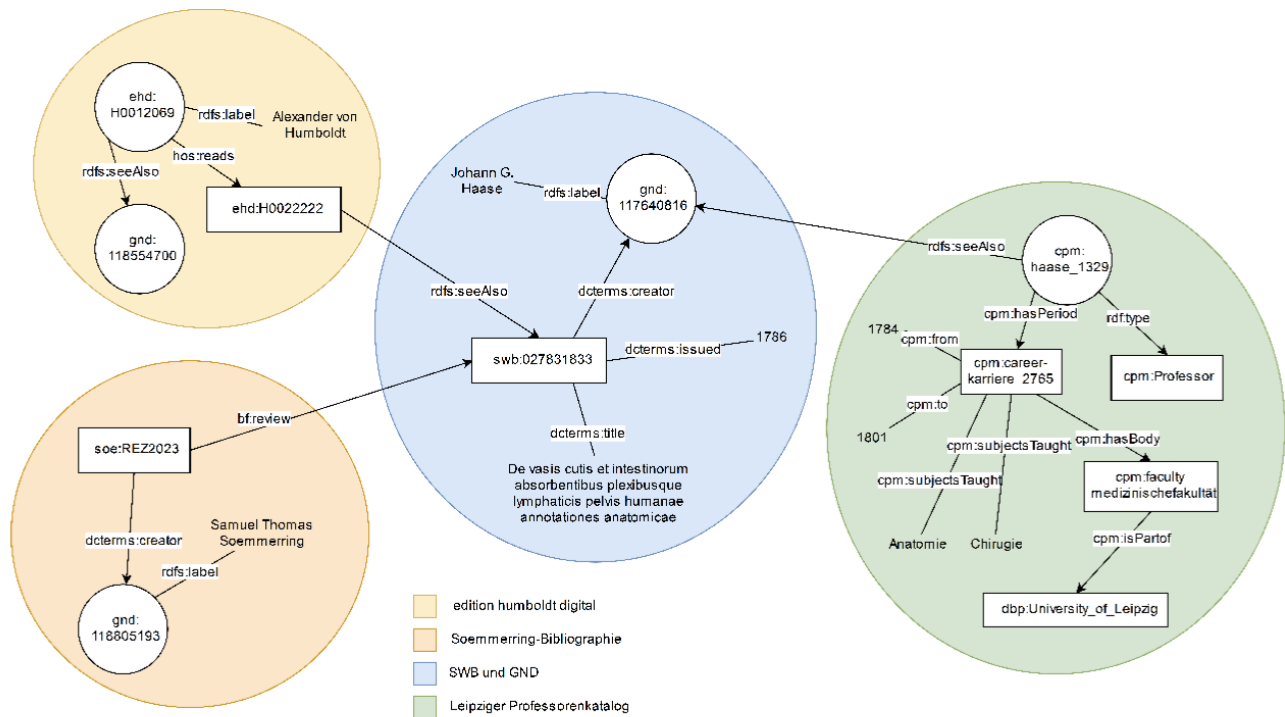


Abb. 2: Semantischer Kommentar aus einer Edition als Teil eines größeren Linked Open Data-Graphen (aus: Dumont, Brief kommentieren im Semantic Web, S. 17)

Informationseinheiten und digitalen Objekte gar nicht erkennbar waren oder nur mit großem Aufwand sichtbar gemacht werden konnten.“⁹³

Dieses Zitat sollte an dieser Stelle deswegen ausführlicher angeführt werden, weil es gleich zwei für unseren Zusammenhang wichtige Aspekte anspricht: Erstens liegt der Reiz des Vernetzungsgedankens nicht nur in der semantisch basierten Verschränkung einzelner Editionen, sondern darüber hinaus auch in der Verbindung mit weiteren Datenpools aus anderen Domänen. Mit Semantic-Web-Technologien ist also die Hoffnung verbunden, das Problem der ‚Datensilos‘, also der isolierten Datenhaltung und -verarbeitung, nicht nur im Rahmen des gesamten Editions wesens, sondern in einem größeren ‚Kulturerbe‘-Maßstab zu überwinden. Zweitens eröffnet diese umfassende Vernetzung nicht nur die Möglichkeit, heterogene Daten über Projekt- und Disziplinengrenzen hinweg einheitlich referenzierbar und ‚ansteuerbar‘ zu machen, sondern auch mit ihnen zu ‚arbeiten‘, also durch die Anreicherung der Editionen mit maschinell auswertbaren Kontexten bisher implizite Informationen zu explizieren und dadurch analysierbar zu machen. Was diese Grundidee des Semantic Web, maschinell mit ‚Bedeutungen‘ zu ‚rechnen‘, für digitale Editionen be-

⁹³ Martin de la Iglesia, Nicolas Moretto und Maximilian Brodhun: Metadaten, LOD und der Mehrwert standardisierter und vernetzter Daten, in: Heike Neuroth, Andrea Rapp, Sibylle Söring (Hgg.), TextGrid: Von der Community – für die Community, Glückstadt 2015, S. 91–104, hier S. 101.

deuten könnte, hat unlängst Stefan Dumont an einem konkreten Beispiel illustriert⁹⁴: Bei Abbildung 2 handelt es sich um die Visualisierung eines Ausschnitts [!] des – in dieser Form fiktiven – Beziehungsnetzes zwischen der *edition humboldt digital* und dem Linked-Open-Data-Graphen. Der Bereich der ersteren ist befindet sich links oben (orange unterlegt). Das Zentrum der Darstellung bildet ein Tripel, das den Sachverhalt „Alexander von Humboldt liest *De vasis cutis*“ durch die beiden projektinternen Identifikatoren ehd: H0012069 (Humboldt) und ehd: H0022222 (*De vasis cutis*) sowie los:reads repräsentiert; letzteres Element ist einer – fiktiven – Ontologie zur Formalisierung wissenschaftshistorischer Kommunikation entnommen. Die Entität „Alexander von Humboldt“ ist mit der Property rdfs:seeAlso mit dem entsprechenden Normdatensatz aus der GND verbunden. Schon hier ergibt sich ein ‚Brückenkopf‘ in die LOD-Cloud, der aber an dieser Stelle von Dumont nicht weiter verfolgt wird. Die Vernetzung wird stattdessen am Beispiel „*De vasis cutis*“ demonstriert, die, ebenfalls via rdfs:seeAlso, mit der entsprechenden Ressource im Online-Katalog des Südwestdeutschen Bibliotheksverbundes verknüpft ist. Da die Katalogdaten ebenfalls als Linked Open Data zur Verfügung stehen, ergeben sich von dort ganz automatisch weitere Verknüpfungen, und zwar sowohl innerhalb des SWB-Katalogs selbst als auch darüber hinaus in weitere Semantic-Web-fähige ‚Datenräume‘. Auch diese Verbindungen sind in der Abbildung – ansatzweise⁹⁵ – visualisiert: Das Werk „*De vasis cutis*“ ist via dcterms:creator mit seinem Verfasser, dem Mediziner Johann Gottlob Haase bzw. dessen GND-Datensatz, verknüpft, der GND-Datensatz wiederum mit der korrespondierenden Ressource im nächsten ‚Datenraum‘, dem *Leipziger Professorenkatalog*, usw. Sind die notwendigen Voraussetzungen erfüllt – Formulierung auf Basis gängiger Standards wie RDF, geeignete Speicherung der Statements (Triplestores, SPARQL-Endpunkte usw.) –, kann das so entstandene ‚Datenmosaik‘ hochkomplexen Analysen unterzogen werden. Dumont nennt als Beispiele: „Eine Humboldt-Forscherin etwa könnte mit der Fragestellung an die Daten herangehen, welche Literatur Alexander von Humboldt wann und zu welchem Thema gelesen und in seinen Briefen diskutiert hat; eine andere Wissenschaftlerin könnte untersuchen, wie medizinische Werke um 1800 gelesen, rezensiert und diskutiert worden sind; eine dritte Wissenschaftlerin, die sich mit der Geschichte der Universität Leipzig befasst, könnte fragen,

⁹⁴ Entnommen aus Dumont, Brief kommentieren im Semantic Web, S. 17-19.

⁹⁵ Die Abbildung stellt auch hier nur einen kleinen Teil des eigentlich Möglichen dar, denn natürlich ergeben sich beispielsweise vom GND-Datensatz Haases noch bedeutend mehr Verbindungen als nur zum Katalogeintrag von „*De vasis cutis*“, zum Leipziger Professorenkatalog und zu rdfs:label „Johann G. Haase“.

wie groß der Einfluss der medizinischen Fakultät Ende des 18. und Anfang des 19. Jahrhunderts gewesen ist. Dies alles auf Grundlage derselben Daten.“⁹⁶

Dass damit der Rahmen des – theoretisch – Möglichen noch nicht erschöpft ist, mag ein weiteres Beispiel veranschaulichen: Bereits 2011 beschäftigte sich Robert Kummer mit Anwendungsszenarien von Semantic-Web-Technologien in der Kodikologie.⁹⁷ Im Mittelpunkt von Kummers Text stand die Vision eines Datennetzes, das RDF-basierte Handschriftenbeschreibungen mit Ressourcen aus bereits im Web vorhandenen Triplestores koppelt. Die auf diese Weise entstandene Kombination von formalisiertem kodikologischen Wissen und Web-Daten würde, so Kummer, nicht nur quantitativ eine neue Dimension für die Kodikologie darstellen, sondern darüber hinaus auch Suchabfragen von einer ungekannten Komplexität ermöglichen. Als Beispiel dafür nannte Kummer die Abfrage: „For the geographic area of Northern Germany, show all codices that contain texts of Classic Latin authors and that have been written in the 13th century. Draw the result as circles on a map and use different colors for monasteries and nunneries.“⁹⁸

Um auf die dieser Arbeit zugrundeliegende Matrix zurückzukommen: Während die eingangs erwähnte bloße Verwendung von Normdaten als eindeutige Identifikatoren auf der ersten Stufe der semantischen Expressivität zu verorten ist, ist die gerade skizzierte Vision einer digitalen Edition, die in ein Netz aus mitunter sehr heterogenen, aber flexibel analysierbaren Daten eingebettet ist, der zweiten und der dritten Stufe der semantischen Expressivität zuzurechnen. Die Grenze zwischen den beiden Stufen ist durch die Frage markiert, ob sich die jeweiligen Editionen darauf beschränken, die Entitäten ihrer Texte nur mit der LOD-Cloud zu verbinden (Stufe 2), oder ob sie darüber hinaus auch Beziehungen zwischen ihnen modellieren (Stufe 3). In ersterem Fall sind die Entitäten gewissermaßen ‚Endpunkte‘ eines größeren Datennetzes. Wie im vierten Abschnitt dieser Arbeit noch zu zeigen sein wird, beschränken sich einige digitale Editionen auf diesen Ansatz, der letzten Endes darauf hinausläuft, dass im Text erwähnte Personen, Orte, Werke und andere Entitäten „zwar als RDF abfragbar [sind], aber ihre Aussagekraft [...] nicht über die Tatsache der Erwähnung [...] hinaus[geht].“⁹⁹ Im Unterschied dazu fußen Inhaltserschließungen, die sich der Stufe 3 zurechnen lassen, darauf, dass sie mithilfe spezifischer Formalisierungen

⁹⁶ Vgl. Dumont, Brief kommentieren im Semantic Web, S. 18.

⁹⁷ Kummer, Semantic Technologies for Manuscript Description.

⁹⁸ Kummer, Semantic Technologies for Manuscript Description, S. 148.

⁹⁹ Dumont, Briefe kommentieren im Semantic Web, S. 3.

Aussagen über die Zusammenhänge der erwähnten Entitäten untereinander treffen. Mit anderen Worten: Auf Stufe 2 bezieht die Edition ihre semantische Expressivität allein aus der Verbindung ihrer Subjekts- und Objektsentitäten mit der LOD-Cloud, auf Stufe 3 darüber hinaus aus einem qualitativen Verhältnis ihrer Entitäten zueinander.

Projekte, die die Kriterien der Stufe 3 in Form einer Ontologie umgesetzt haben, sind mittlerweile vergleichsweise älteren Datums: Mit den bereits erwähnten *Henry III Fine Rolls* hat schon eines der ersten Editionsprojekte zur Umsetzung von Semantic-Web-Technologien auf diesen Ansatz zurückgegriffen.¹⁰⁰ Ziel des Projekts war die digitale Edition der Finanzaufzeichnungen, der sogenannten fine rolls, der Kanzlei des englischen Königs Henry III. Die fine rolls dokumentieren einen typischen Akt mittelalterlicher Herrschaftsausübung, nämlich die Vergabe von königlichen Ämtern und Privilegien gegen Bezahlung. Um diesen Vorgang abzubilden, wurde – im Wesentlichen unter Verwendung von OWL – eine Ontologie entwickelt, die bereits verbreitete Standards (CIDOC CRM, Dublin Core sowie W3-Proposals wie Geo, SKOS, Time usw.) mit projektspezifischen Elementen kombiniert.¹⁰¹ Zu Letzteren zählt beispielsweise eine eigene Klasse für verwandtschaftliche Verhältnisse (`frh3:Relationship`) oder die Property `frh3:toponyme_name`, mit der sich die bei mittelalterlichen Namen so häufige toponymische Personenbezeichnung als Triple aus Vornamen, besagter Property und Ort ausdrücken lässt.¹⁰²

Das Ziel dieses Ansatzes besteht darin, durch die Zerlegung des Textes in seine semantischen ‚Bausteine‘ angereicherte Indizes für Orte, Personen, Verwandtschaftsverhältnisse und „Rollen“ (*roles*)¹⁰³ zu generieren. Auf diese Weise sollen komplexe, aber ursprünglich verdeckte Beziehungen zwischen den im Text erwähnten Entitäten aufgezeigt werden, um Einblicke in vielschichtige Phänomene wie mittelalterliche Karrieremuster, genealogische Verhältnisse und geografische Zusammenhänge zu gewinnen. Vor diesem Hintergrund wird auch deutlich, dass Properties wie `frh3:toponyme_name` keine technischen Spielereien darstellen, sondern mit Blick auf die mediävistische Forschung entwickelte Instru-

¹⁰⁰ Vgl. Ciula et al.: Implementing an RDF/OWL Ontology on Henry the III Fine Rolls.

¹⁰¹ Zur technischen Seite des Projekts vgl. v.a. dies. et al.: Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project.

¹⁰² Beispielsweise würde sich Thomas von Aquin zerlegen lassen in „Thomas“ – `frh3:toponyme_name` – „Aquino“, Bernhard von Clairvaux in „Bernhard“ – `frh3:toponyme_name` – „Clairvaux“ usw.

¹⁰³ Unter Letzterem ist das Verhältnis zu verstehen, in dem die erwähnten Personen zueinander stehen. So enthält die Klasse `Role_Relationship` etwa die Unterklasse `King_Servant`, mit der wiederum die Properties `frh3:has_master` und `frh3:has_servant` verbunden sind; Beispiel nach Ciula et al., Implementing an RDF/OWL Ontology on Henry the III Fine Rolls, S. 7.

mente sind, um Spezifika der Dokumente – hier den ‚geographische Gehalt‘ mittelalterlicher Namen – für Semantic-Web-Technologien operationalisierbar zu machen.

Neben den *Henry III Fine Rolls* mit ihrer Grundidee eines erweiterten Index wurden in den letzten Jahren weitere Ansätze zum Einsatz von Ontologien für die inhaltliche Erschließung historischer Materialien entwickelt. Besondere Aufmerksamkeit hat dabei die Bookkeeping Ontology, ein von Georg Vogeler und seinen Mitarbeitern am Zentrum für Informationsmodellierung der Universität Graz für die Erschließung von Rechnungsschriftgut entwickeltes formales Modell, erfahren.¹⁰⁴ Für Vogeler bietet die Entwicklung eines komplexen Datenmodells wie der Bookkeeping Ontology einen entscheidenden Mehrwert für die digitale Edition von wirtschaftshistorisch relevantem Material. Denn bisher, so Vogeler, bedienten sich Editionen von historischen Rechnungsdokumenten entweder der XML/TEI-Standards, um paläographische oder linguistische Aspekte auszuzeichnen, oder sie fokussierten sich auf die Extraktion ökonomischer Daten in Form basaler buchhalterischer Fakten, die entweder auf simple Tabellen übertragen oder mithilfe der – ebenfalls auf XML beruhenden – eXtensible Business Reporting Language (XBRL)¹⁰⁵ kodiert werden. Eine Entscheidung für einen Aspekt bedeutet damit immer auch eine Einengung auf ein Forschungsinteresse, unter dessen Maßgabe der Text ediert wird: Entweder visuell (paläographisch) oder linguistisch (sprachhistorisch) oder inhaltlich (wirtschaftshistorisch). Anknüpfend an Vorüberlegungen von Manfred Thaller¹⁰⁶, schlägt Vogeler daher alternativ die Erprobung von Semantic-Web-Technologien bei der Edition von Rechnungsschriftgut vor. Kernstück dieses Vorhabens ist die bereits erwähnte Bookkeeping Ontology,¹⁰⁷ die gewissermaßen als formalisiertes Grundgerüst ökonomischer Austauschbeziehungen fungieren soll: *Dass* ein ökonomischer Tauschakt stattgefunden hat (bk:Transaction); dass daran zwei *Seiten* teilgenommen haben (bk:Party bzw. bk:Account); dass dabei *messbare Größen* (bk:Measurable) ausgetauscht wurden, durch die *Verbindlichkeiten* (bk:Liabilities) entstehen können; dass dieser Vorgang an einem *Ort* (bk:Place) vor

¹⁰⁴ Für die folgenden Ausführungen vgl. v.a. Vogeler, *The Contents of Accounts*; ders., Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?

¹⁰⁵ Die eXtensible Business Reporting Language ist eine W3-Empfehlung, die hauptsächlich zur Standardisierung von Finanzdaten entwickelt wurde; vgl. <https://de.xbrl.org>.

¹⁰⁶ Vgl. Manfred Thaller: *What is a text within the Digital Humanities, or some of them, at least?* In: *digital humanities 2012*. Hamburg 2012. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/beyond-embedded-markup/>.

¹⁰⁷ <https://github.com/GVogeler/bookkeeping>

sich ging und von einer *Quelle bezeugt* (bk:attestet) werden kann usw.¹⁰⁸ Die Bookkeeping Ontology stellt also mit anderen Worten ein Vokabular dar, auf dessen Konzepte sich alle wirtschaftlichen Transaktionen reduzieren lassen sollen. Ihr Ziel besteht letztlich darin, durch die Bereitstellung einer einheitlichen formalen Folie alle in historischem Rechnungsschriftgut auftretenden „Rechnungsentitäten“¹⁰⁹ zu aggregieren und miteinander in Beziehung zu setzen.

Allerdings gibt Vogeler selbst zu bedenken, dass die Bookkeeping Ontology dafür einer entscheidenden Erweiterung bedarf, da sie nur die Struktur der wirtschaftlichen Handlung, nicht aber deren konkreten Inhalte bzw. „values“ (also v.a. Art und Umfang der Geldbeträge und Waren) abbildet. Um dieses Problem zu lösen, regt Vogeler zusätzlich die Erstellung thematischer Taxonomien mit SKOS an. Mit diesen ließen sich verschiedene Währungs- und Warensysteme, -einheiten und -bezeichnungen, aber auch Maßeinheiten und sogar Berufsgruppen zu Taxonomien zusammenfassen. Auf dieser Grundlage ließen sich, so Vogeler, so unterschiedliche Korpora wie frühneuzeitliche dänische Zollregister und Baseler Stadtrechnungen des 17. Jahrhunderts miteinander in einen maschinenprozessierbaren Zusammenhang setzen und – theoretisch – um immer neue Quellenbestände unterschiedlicher Provenienz erweitern.

Auch die Bookkeeping Ontology lässt sich auf der von mir eingangs vorgestellten Matrix verorten. Da sie für die Erschließung von Rechnungsschriftgut, also hoch strukturiertem Quellenmaterial, gedacht ist und auf einer Kombination aus semantisch expressivem formalen Modell (Expressivitätsstufe 3) und fortgeschrittener Abstraktionsstufe (konkrete und abstrakte Entitäten) fußt, ist sie ein paradigmatischer Vertreter des Matrixfeldes 6.

Damit wurden bis zu diesem Punkt verschiedene praktische und theoretische Ansätze vorgestellt, die zwar alle Grade der semantischen Expressivität abgedeckt haben, auf der Ebene der Abstraktion aber auf die Stufen 1 und 2 beschränken. Im Folgenden sollen nun Überlegungen vorgestellt werden, deren Ansatz darin besteht, sowohl konkrete als auch abstrakte Entitäten von komplexen, unstrukturierten Texten in einem formalen Modell zu erschließen (Matrixfeld 9).

Der erste Ansatz, auf den ich in diesem Zusammenhang verweisen möchte, ist die von Stefan Heßbrüggen-Walter im Rahmen zweier Aufsätze entwickelte Minimal Doxographi-

¹⁰⁸ Beispiele nach Vogeler, *The Contents of Accounts*, S. 9

¹⁰⁹ So die von Georg Vogeler selbst in meinen Fragebogen gebrauchte Bezeichnung.

cal Ontology.¹¹⁰ Den Ausgangspunkt seiner Überlegungen bildet die Unterscheidung von drei verschiedenen Formen von Tatsachen/Fakten (facts):

1.) Historische Fakten. Unter diesen sind alle ‚realen‘ Entitäten (bzw. „spatiotemporal entities“) zu verstehen, also in erster Linie Philosophen und konkrete Werke, aber auch Ereignisse.

2.) Philosophische Fakten. Hiermit sind weniger objektive Tatsachen als alle denkbaren philosophische Aussagen („Raum und Zeit sind notwendige Vorstellungen“; „Gott ist tot“; „Alle Philosophie ist ‚Sprachkritik‘“; „Es gibt kein richtiges Leben im falschen [Leben]“ usw.) gemeint. Als Philosophisches Faktum kann also „jeder Behauptungsinhalt gelten, sofern er einmal in einem philosophischen Text formuliert worden ist.“¹¹¹ Werden solche Philosophischen Fakten in Form eines Tripels formuliert, ergeben sich zwei zentrale Unterschiede zu Historischen Fakten. Zum ersten muss ein „Behauptungsinhalt“ keineswegs sachlich richtig sein, um als (Philosophisches) Faktum zu gelten. In diesem Sinne ist die Behauptung, dass Raum und Zeit notwendige Vorstellungen sind, allein durch den Umstand, dass sie formuliert worden ist, eine gültige Aussage im Sinne eines Philosophischen Faktums. Im Gegensatz dazu gilt die Aussage, dass Immanuel Kant die *Kritik der reinen Vernunft* verfasst hat, nur dann als (Historisches) Faktum, wenn sie als objektiv zutreffend erkannt ist. Der zweite Unterschied besteht darin, dass bei als Tripel formulierten Philosophischen Fakten keine ‚realen‘, sondern ‚abstrakte‘ Entitäten an der Subjekt- und der Objektposition stehen müssen. Unter ‚abstrakten‘ Entitäten können verschiedenste philosophiehistorisch relevante Konzepte verstanden werden. Dazu können neben Nahelegendem wie „Gott“, „Geist“ und „Idee“ usw. auch spezifische Grundelemente spezieller doxographischer Diskurse wie „Befreiung der Seele“, „Tierische Vernunft“ oder „Aus Idee abgeleitete Regel“ [!] zählen.¹¹²

3.) Verbindende Fakten. Die Verbindenden Fakten stellen den Zusammenhang zwischen dem ‚historischen‘ und dem ‚philosophischen‘ Bereich her, indem sie das Verhältnis der ‚realen‘ Entitäten zu den Philosophischen Fakten genauer definieren. Heßbrüggen-Walther beschränkt sich an dieser Stelle auf die drei grundlegenden Bezugskategorien Zustim-

¹¹⁰ Vgl. Heßbrüggen-Walter: Tatsachen im semantischen Web: Nanopublikationen in den digitalen Geisteswissenschaften?, in: Peter Haber, Eva Pfanzelter (Hgg.), *Historyblogosphere. Bloggen in den Geschichtswissenschaften*, München 2013, S. 149–160. DOI: 10.1524/9783486755732.149; ders., *What People Said*.

¹¹¹ Heßbrüggen-Walter: *Tatsachen im semantischen Web*, S. 155.

¹¹² Letzteres Beispiel nach ebd.

mung (asserts/accepts), Ablehnung (denies) und (neutrale) Beschäftigung (reflects). Schon diese bewusst limitierte Anzahl von Operatoren ermöglicht die Formulierung von komplexen vollständigen doxographischen Tatsachen oder – um an die Diktion Heßbrüggen-Walthers anzuknüpfen – propositionalen Strukturen: Der ‚propositionale Träger‘ (also die ‚reale‘ Entität eines Philosophen oder eines Werks) artikuliert seine ‚propositionale Einstellung‘ (also Zustimmung, Ablehnung, Reflexion) bezüglich eines ‚propositionalen Gehalts‘ (also eines Philosophischen Faktums in seiner Gesamtheit).

Auf technisch-formaler Ebene lässt sich dieses Muster von Aussagen über Aussagen als Reifikation fassen. Um auf das obige Beispiel zurückzugreifen: Gegeben sei die doxographische Tatsache, dass Immanuel Kant als Person (oder auch die *Kritik der reinen Vernunft* als Werk) behauptet, dass Raum und Zeit notwendige Vorstellungen sind. Letzteres stellt ein Philosophisches Faktum dar, das sich als Statement aus „Raum und Zeit“ (Subjekt), „ist“ (Prädikat) und „notwendige Vorstellung“ (Objekt) formalisieren ließe. Dieses Statement kann nun wiederum als Objekt eines weiteren Statements fungieren, dessen Subjekt die ‚reale‘ Entität Immanuel Kant und dessen Prädikat die Einstellung ‚Zustimmung‘ ist.

Da diese Struktur aus (philosophischer) Aussage in Relation zu einem ‚Träger‘ dem Grundprinzip doxographischer Diskurse (welche Aussagen gibt es und wer hat welches grundlegende Verhältnis zu ihnen) entspricht, hält sie Heßbrüggen-Walter auch für ausreichend, um so komplexe Themenfelder wie philosophiehistorische Diskussionen in Semantic-Web-fähiger Form abzubilden und analysierbar zu machen. Als Proof of Concept hat Heßbrüggen-Walter selbst einige ausgewählte frühneuzeitliche philosophische Texte auf dieser Grundlage erschlossen und in Form von Nanopublikationen veröffentlicht.¹¹³ Er betont zugleich, dass dieser Ansatz bewusst nicht auf die semantische Bindekraft einer schwergewichtigen Ontologie baue und eben nur eine „*minimal doxographical ontology*“¹¹⁴ darstelle. Sie habe das Ziel, Diskurse abzubilden „without making any assumptions about the conceptual structure of the respective domain.“¹¹⁵ Diesen Verzicht auf umfassendere formale Spezifikationen will Heßbrüggen-Walter eher als Vorteil verstanden wissen: „The main advantage of such an approach over proper reification is that it can be used heuristically: we do not need a full blown ontology for capturing the content of a given discourse in

¹¹³ http://emto-nanopub.referata.com/wiki/EMTO_Nanopub [26.04.2019]

¹¹⁴ Hervorhebung Till Meyer.

¹¹⁵ Heßbrüggen-Walter, What people said, S. [9].

a form that is amenable to further refinement and development.“¹¹⁶ Es sei zwar prinzipiell möglich, die ‚doxographischen Tripel‘ in ausgewachsene („full blown“) Ontologien einzubinden. Letztere müssten aber immer im Hinblick auf konkrete philosophiehistorische oder andere fachspezifische Konstellationen entworfen werden. Damit würden auch die ‚doxographischen Fakten‘ ihren Charakter als möglichst neutrale Repräsentationen von Meinungen verlieren: „By transforming the isolated statements of a doxographical record into the reconstruction of a conceptual scheme, we leave the domain of ›facts‹ and enter into the realm of interpretation.“¹¹⁷

Die Einordnung des Ansatzes von Heßbrüggen-Walter auf der Matrix des Einleitungsteils dieser Arbeit fällt nicht ganz leicht: Der Abstraktionsgrad der Entitäten ist zweifelsohne auf der höchsten Stufe (Stufe 3) anzusiedeln. Heßbrüggen-Walters bewusste Zurückhaltung bei der formalen Spezifikation macht allerdings eine Differenzierung notwendig: Seine „minimal doxographical ontology“ ist zwar formal der dritten Stufe der semantischen Expressivität zuzuordnen, allerdings bleibt sie hinter den Möglichkeiten einer schwergewichtigen Ontologie zurück.

Letztere voll auszuschöpfen ist die Motivation der Überlegungen Jochen Strobels, der am Beispiel der Korrespondenz der Jenaer Frühromantiker die Möglichkeiten einer auf Konzepten der Frame-Semantik aufbauenden Inhaltsanalyse beschreibt.¹¹⁸ Unter Frames werden im linguistischen Sinn erfahrungsbasierte Konzepte verstanden, mit denen Wissensausschnitte schematisch organisiert und in Beziehung zu einem größeren Weltwissen gesetzt werden. Jeder Frame oder Wissensrahmen wird durch ein Ensemble von Wissenselementen, sogenannten Slots, definiert. Strobel selbst illustriert dies an einem Beispiel: Mit dem Frame „Kindergeburtstag“ sind bestimmte Slots (z.B. Getränke, Spiele, bestimmte Orte usw.) verbunden, die wiederum mit bestimmten Werten (Fanta, Sackhüpfen, Kinderzimmer usw.) befüllt sind.¹¹⁹ Bei Frames handelt es sich also mit anderen Worten um begriffsbasierte Repräsentationen kollektiver Wissenskonzepte.

Strobels Vorschlag läuft nun darauf hinaus, Frames und die sie konstituierenden Slots zur Bildung von Begriffssystemen zu nutzen, in die nichts weniger als „generelle[s] Wissen vom Menschen und den kognitiven und emotionalen Vorgängen in ihm, allgemeines

¹¹⁶ Ebd., S. [8].

¹¹⁷ Ebd., S. [12].

¹¹⁸ Vgl. Strobel, Digitale Briefedition und semantische Erschließung.

¹¹⁹ Vgl. Ebd., S. 158f.

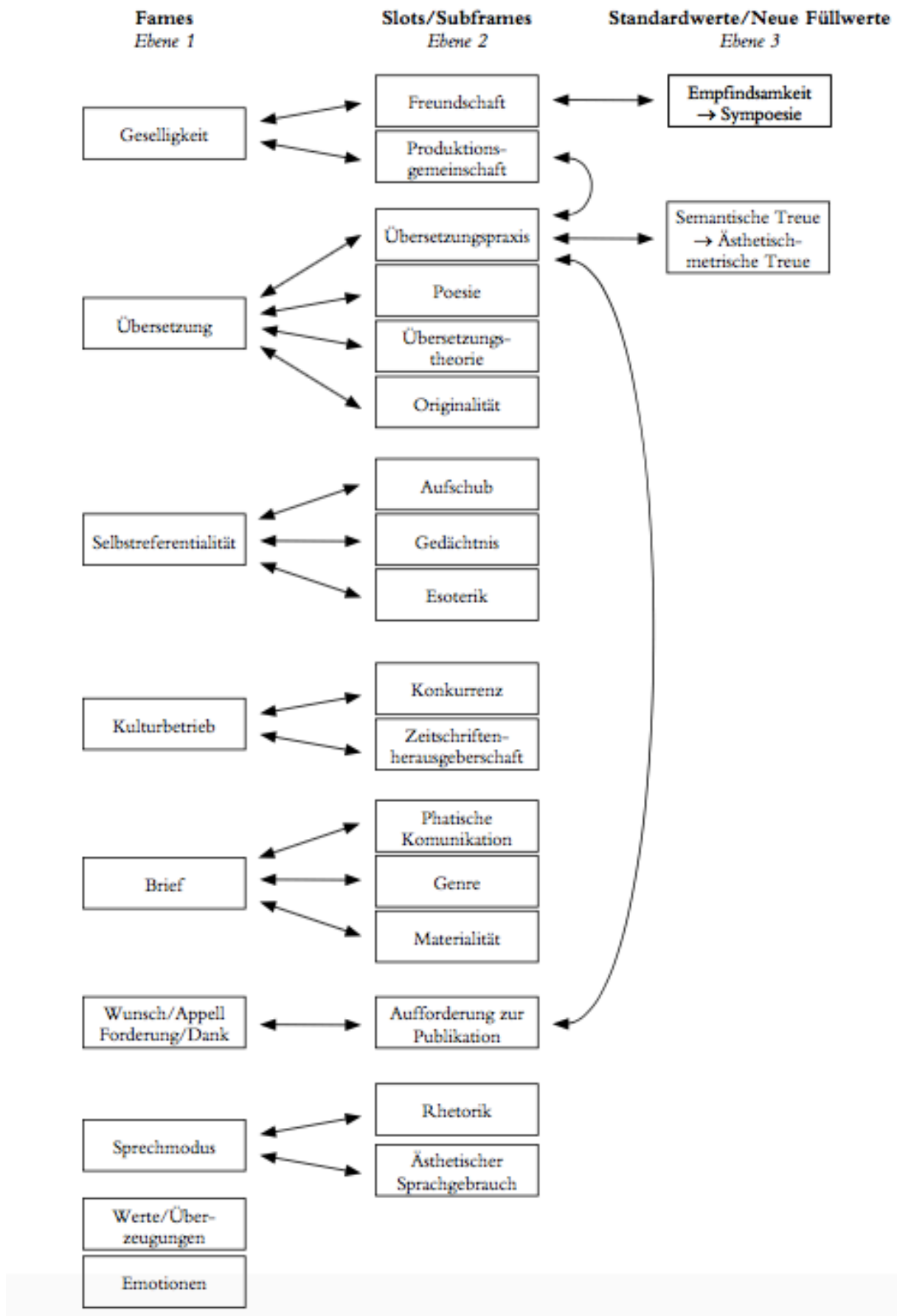


Abb. 3: Semantische Erschließung 'Romantikerbrief'. Exemplarische Frames (aus Strobel, *Digitale Briefedition und semantische Erschließung*, S. 173)

Weltwissen, spezielles zeitgenössisches für die Entstehungszeit [...] und noch spezieller das gruppeninterne Wissen der Korrespondenzpartner“¹²⁰ einfließen sollen. Je nach Quellenkorpus und Forschungsinteresse wären dafür in einem ersten Schritt zentrale Frames zu definieren, was in diesem Fall bedeutet: Wissenskonzepte im Hinblick auf ihren (historischen) Kontext (insbesondere Zeit und Personenkreis) zu rekonstruieren. Dies bedeutet auch, die dazugehörigen Slots und Werte hierarchisch zusammenzufassen, wobei die Werte ihren Slots, die Slots wiederum dem Frame untergeordnet wären. Am obigen Beispiel illustriert ergäbe sich also die Taxonomie mit der Oberkategorie „Kindergeburtstag“, dem u.a. die Kategorie „Getränke“ untergeordnet wäre, unter der wiederum Kategorien wie „Fanta“ angesiedelt wäre. Durch Querverbindungen zwischen einzelnen Elementen verschiedener Taxonomien würde sich eine komplexe Ontologie ergeben. Denkbar wäre z.B., um beim obigen Beispiel zu bleiben, die Ergänzungen des „Kindergeburtstags“ um einen Frame wie „Markenprodukte“, in dessen Kategorie „Nahrungsmittel“ ein Begriff wie „Fanta“ aufgeführt sein kann. Dabei lässt sich die Auswahl der Frames sehr flexibel handhaben, wie ein Blick auf die von Strobel vorgestellten Beispiele (vgl. Abb. 3) verdeutlicht: Zur inhaltlichen Analyse der Korrespondenz der Jenaer Frühromantiker schlägt er so verschiedene und bisweilen äußerst abstrakte Frames wie „Geselligkeit“, „Übersetzung“, „Selbstreferentialität“ oder „Wunsch/Appell/Forderung/Dank“ vor.

Idealerweise würde sich auf Grundlage einer solchen Frame-Analyse ein strukturiertes Begriffsnetz ergeben, das alle relevanten Themen eines spezifischen Quellenkorpus abdeckt. Strobel sieht ein ganz besonderen Mehrwert einer solchen Vorgehensweise darin, dass eine solche Ontologie zeit- und gruppenspezifische Zusammenhänge sichtbar macht und es so ermöglicht, „Texte auf eine Forschungsfrage hin zu erschließen, dabei auch Kontextwissen (wie Entstehungsbedingungen und Wirkung) heranzuziehen und Inferenzen vorzunehmen, also für das Verständnis wichtige, aber nicht explizite Inhalte gleich mit zu erschließen – Ergebnis ist ein später statistisch auswertbares Datenmodell, das wiederum auf Selektion, Abstraktion und Klassifikation beruht.“¹² In dieses Modell fließt die Summe des dem Materialkorpus zu entnehmenden Wissens ein, das sich zu der gegebenen Forschungsfrage in Beziehung setzen lässt.“¹²¹

Abschließend bleibt festzuhalten, dass Strobels Überlegungen – zumindest unter den Vorzeichen des ‚konzeptuellen‘ Zuschnitts der vorliegenden Arbeit – das elaborierteste,

¹²⁰ Ebd., S. 159.

¹²¹ Ebd., S. 156.

den maximalen Abstraktionsgrad mit maximaler semantischer Expressivität vereinende Konzept darstellen.

4. Semantic-Web-Technologien & digitale Editionen in der Praxis

Von den von mir untersuchten 52 Editionen ließen sich fünf keinem Matrixfeld zuordnen, da sie keine Normdaten, geschweige denn Semantic-Web-Technologien verwenden.¹²² Von den verbliebenen 47 Editionen lassen sich 30 dem Expressivitätsgrad 1 zuordnen (vgl. hier und im Folgenden die Matrix in Abb. 4).¹²³ Diese Editionen verwenden also Normdaten wie GND- oder Ge-Names-Datensätze, allerdings nur zum Zweck der eindeutigen Referenzierung. Von den 30 Editionen beschränken sich 23 darauf, konkrete Entitäten auszuzeichnen. Die Mehrheit der Projekte dieses Expressivitätsgrades lässt sich also Feld 1 und damit dem ‚einfachsten‘ Konzept der Matrix zuordnen. Die restlichen 7 Editionen verwenden in der einen oder anderen Form Schlagwörter und dehnen damit ihre Erschließungstiefe auch auf abstrakte Konzepte aus. Eine Edition (*Augsburger Baumeisterbücher*) lässt sich als Rechnungsschriftgut Matrixfeld 4 zuordnen, bei allen anderen Projekten handelt es sich um Editionen von Briefen (*Digitale Edition der Briefe Erdmuthe Benignas von Reuß-Ebersdorf*; *Alfred Escher-Briefedition*; *Vernetzte Korrespondenzen: Exilbriefnetz*; *Aloys Hirt – Briefwechsel*), Tagebüchern (*Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg*) oder Reisebeschreibungen (*Die Reisen des Botanikers Carl Haussknecht*), also Beispiele für das Matrixfeld 7.

Nach Abzug der 30 Projekte der Matrixfelder 1, 4 und 7 (Expressivitätsgrad 1) bleiben 17 Editionen, die sich den übrigen Matrixfeldern zuordnen lassen und damit im eigentlichen Sinne Semantic-Web-Technologien anwenden.

Mit insgesamt acht Vertretern ist hier das Matrixfeld 2 am stärksten vertreten. Dieses ist dadurch definiert, dass die konkreten Entitäten der Edition – erstens – als RDF über SPARQL abfragbar und damit Teil des Semantic Web sind, dass aber – zweitens – im

¹²² Konkret handelt es sich um die Projekte *Frühneuzeitliche Fürstinnenkorrespondenz im mitteleuropäischen Raum*, *Die Kabinettsprotokolle der Landesregierung von Nordrhein-Westfalen*, *Erich Mühsam - Tagebücher*, *Kommentierte Edition der Briefe Ernst Tollers* und *Edition der Korrespondenz Frank Wedekinds als Online-Volltextdatenbank*.

¹²³ Vor Schwierigkeiten stellte mich die Edition *Blumenbach – Online*: Das Projekt als solches ist zwar – soweit aus der Projektbeschreibung und der Sekundärliteratur (vgl. Wettläufer et al., *Semantic Blumenbach*) ersichtlich – eindeutig dem Matrixfeld 1 zuzuordnen. Allerdings wurde in Zusammenhang mit *Blumenbach – Online* mit *Semantic Blumenbach* ein kleines Seitenprojekt durchgeführt, dessen Ziel die Verbindung von Bild und Text auf Basis von CIDOC CRM ist. Das Vorgehen bestand aber letztlich ‚nur‘ darin, mithilfe der Property `ecrm:P48_has-preferred-identifier` physische Sammlungsobjekte der Blumenbach-Sammlung und deren TEI-kodierte Erwähnung im Text der Blumenbach-Edition zu einem gemeinsamen projektinternen Kerndaten-Identifizier zusammenzuführen. Da sich ein solches Vorgehen auf meiner Matrix erstens schwer verorten lässt und es sich zudem in Wesentlichen um ein Testprojekt im Rahmen des ‚eigentlichen‘ Projekts *Blumenbach – Online* handelt, hat *Semantic Blumenbach* bei der Auswertung keine weitere Berücksichtigung gefunden.

	Expressivitätsgrad 1: Normdaten zur Referenzierung	Expressivitätsgrad 2: Normdaten als LOD	Expressivitätsgrad 3: Modellierung interner Bezüge
Abstraktionsgrad 1: konkrete Entitäten, stark und schwach strukturierte Texte	23	8 (Burckhardtsource; Theodor Fontane: Notizbücher; Carl-Maria-von-Weber- Gesamtausgabe; <i>Austrian precursors</i> ; Joseph Eckhel; Uwe Johnson- Werkausgabe; Pacelli: <i>Nuntiaturberichte</i> ; Arthur Schnitzler – <i>Briefwechsel mit Autorinnen</i>)	3 (Sandart.net; Bibliothek der Neologie; <i>Der Sturm</i>) Dumont: Briefe annotieren im Semantic Web
Abstraktionsgrad 2: konkrete und abstrakte Entitäten, stark strukturierte Texte	1	0	3 (Baseler Urfehdebücher; Baseler Jahrrechnungen; DEPCHA) u.a. Vogeler: <i>The Content of Accounts...</i>
Abstraktionsgrad 3: konkrete und abstrakte Entitäten, schwach strukturierte Texte	6	1 (Lavater: <i>Briefwechsel</i>)	[2] [<i>Der Welsche Gast</i>]; [Stefan George Digital] Heßbrüggen-Walter: <i>What People said</i> ; Strobel: <i>Digitale Briefedition..</i>

Abb. 4: Abstraktions-/Expressivitätsmatrix (Projekte, bei den die Arbeit am Datenmodell nicht nicht abgeschlossen sind, sind kursiv gekennzeichnet; Sekundärliteratur für die Matrixfelder ist grau gekennzeichnet)

Rahmen der Edition keine Bezüge zwischen ihnen modelliert sind. Allerdings ist zu bedenken, dass nur drei Projekte (*Burckhardtsource*, *Carl-Maria-von-Weber-Gesamtausgabe*, *Theodor Fontane: Notizbücher. Digitale genetisch-kritische und kommentierte Edition*) diese Struktur auch tatsächlich schon in der Praxis anwenden. Für die restlichen fünf Projekte ist laut Fragebögen eine Umsetzung, also vor allem die Bereitstellung der Daten als RDF, zwar in Planung oder in Arbeit, aber noch nicht abgeschlossen. Ähnlich gelagert ist die Situation bei *Lavater: Historisch-kritischen Edition ausgewählter Briefwechsel*, dem einzigen Projekt, das ich dem Matrixfeld 8 zuordnen würde: Die Bereitstellung der Editionsdaten als RDF ist über die Nationale Infrastruktur für Editionen NIE – INE¹²⁴ zwar anvisiert, aber noch nicht umgesetzt.

¹²⁴ <https://www.nie-ine.ch> [15.05.2019]

Als ein Vertreter des Matrixfeldes 2 soll an dieser Stelle das Projekt *Burckhardtsource* näher vorgestellt werden.¹²⁵ Ziel des Projektes, das seit 2010 unter der Federführung der Scuola Normale Superiore Pisa/Florenz durchgeführt wird, ist die kritische Online-Edition von rund 1100 größtenteils unveröffentlichten Briefen von 400 unterschiedlichen Korrespondenzpartnern an den berühmten Schweizer Kulturhistoriker Jacob Burckhardt. Diese werden in drei Schichten präsentiert: als Faksimile des Manuskripts, in einer XML/TEI-kodierten Transkription sowie als 'semantisch annotierter' Text bzw. „Semantic Edition“.¹²⁶ Die 'semantische Annotation' wird mithilfe des Open-Source-Tools Pundit¹²⁷ realisiert, mit dem sich beliebige Webseitenbestandteile, also beispielsweise einzelne Wörter oder Abschnitte eines digital edierten Briefs, mit dem Web of Data verbinden lassen. Dafür werden in einem ersten Schritt die Entitäten der Webseite ausgezeichnet, im Falle des Burckhardt-Projekts sind dies die in den Briefen erwähnten Personen, Orte, Kunstwerke und Zeitanangaben. In einem zweiten Schritt werden die ausgezeichneten Entitäten durch Properties mit einer Wissensdatenbank verknüpft; ursprünglich war dies die mittlerweile in Wikidata aufgegangene Datenbank Freebase. Daraus ergibt sich ein Subjekt-Prädikat-Objekt-Statement, das als RDF-Tripel abgespeichert wird. Der Mehrwert besteht also letztlich darin, dass Annotationen in strukturierte semantische Daten umgewandelt und so in das Web of Data integriert werden.

Für die nähere Qualifizierung des Verhältnisses zwischen Web-Ressourcen einerseits und Datenbank-Ressourcen andererseits haben die Entwickler von Pundit einige wenige 'Grundrelationen' in Form von Properties wie „criticizes“, „quotes“, „has topic“ usw. zur Verfügung gestellt. *Burckhardtsource* schöpft diese Möglichkeiten allerdings kaum aus und beschränkt sich darauf, die in den Briefen erwähnten Entitäten durch die Properties „identifies person“ oder „identifies place“ mit den korrespondierenden Datenbankressourcen zu verknüpfen. Doch schon durch diese einfache Operation, die letztlich auf eine eindeutige Identifizierung der Editionsentitäten mit Linked-Open-Data-fähigen Normdatensätzen hinausläuft, wird die Edition mit weiterem Kontextwissen angereichert, das in die Analyse des Materials einbezogen werden kann. So soll es beispielsweise nicht nur möglich sein, alle in einem Brief genannten Orte – als explizite Informationen – auf einer interaktiven Land-

¹²⁵ <https://www.burckhardtsource.org> [08.05.2019]

¹²⁶ So die Bezeichnung des entsprechenden Reiters auf der Projektwebseite in Abhebung von der „Philological Version“, unter der das Faksimile und die 'einfache' Transkription zu finden sind.

¹²⁷ <https://thepundit.it>; speziell zur Arbeit mit Pundit im *Burckhardtsource*-Projekt vgl. Francesca di Donato et al.: Semantic Annotation with Pundit. A case study and a practical demonstration, in: Proceedings of DH-Case 2013, Florence, 2013. DOI: 10.1145/2517978.2517995.

karte anzuzeigen, sondern auch die Geburtsorte der erwähnten Personen – also implizite Informationen – zu visualisieren!¹²⁸ Weitere solcher Nutzungsszenarien auf Basis ‚semantischer Annotationen‘ sind denkbar, etwa eine Landkarte, auf der angezeigt wird, wo sich alle in einem Brief erwähnten Kunstwerke zum Zeitpunkt der Abfassung des Briefs befinden oder dergleichen – insbesondere für Kunsthistoriker relevante – Zusammenhänge.¹²⁹

Ebenfalls in einem kunsthistorischen Zusammenhang kommen Semantic-Web-Technologien bei *Sandrart.net* dem ältesten Projekt, das sich dem Matrixfeld 3 zuordnen lässt, zum Einsatz.¹³⁰ Gegenstand des 2012 abgeschlossenen Projekts war die Edition der „Teutschen Academie der Edlen Bau- Bild- und Mahlerey-Künste“, also des kunstwissenschaftlichen Hauptwerks des Malers, Kupferstechers und barocken Kunsttheoretikers Joachim von Sandrart. Ähnlich wie *Burckhardtsource* wurden dafür die im Text erwähnten Personen, Orte und Kunstwerke sowie Publikationen ausgezeichnet. Was *Sandrart.net* nun von *Burckhardtsource* unterscheidet, sind die Querverweise zwischen den Entitäten innerhalb des Korpus: Die „Teutsche Academie“ ist reich an internen impliziten und expliziten Bezügen (ein Ölbild X, das auf S. 122 erwähnt wird, diente Kupferstecher A, der auf S. 255 erwähnt wird, als Vorlage für Kupferstich Y, der auf S. 312 erwähnt wird; die als Büste von Person A (erwähnt S. 53) befindet sich in Besitz von Person B (erwähnt S. 512), der zugleich Schwiegersohn von A ist; die auf der Radierung Z abgebildete Stadt C ist Entstehungsort der in der „Teutschen Academie“ erwähnten Kunstwerke 1,2 und 3 usw.). Dieses Quellencharakteristikum fand in der Datenmodellierung insofern ihren Niederschlag, als die Entitäten der Edition nicht nur mit externen Referenzdaten, sondern auch untereinander verknüpft wurden. Dies geschieht in Form von gerichteten Graphen, die zu großen Teilen¹³¹ in RDF-Tripel überführt wurden und als Linked Open Data angeboten werden. Auf dieser Grundlage wurde verschiedene intratextuelle Bezüge zwischen den Entitäten der erwähnten Kategorien (Personen, Orte, Kunstwerke, Publikationen) modelliert. Eine vollständige Übersicht über die Elemente des Datenmodells von *Sandrart.net* lässt sich zwar weder aus der Projektwebseite noch aus der einschlägigen Literatur gewinnen, die Angaben des dankenswerterweise ausgefüllten Fragebogens sowie eine gezielte Recherche

¹²⁸ Vgl. Donato et al.: Semantic Annotation with Pundit, S. 4.

¹²⁹ Vgl. ebd., S. 2.

¹³⁰ www.sandrart.net/de/ [09.05.2019]

¹³¹ Laut Projektwebseite war „eine sinnvolle ›Übersetzung‹ unserer Bezugs-Tripel in RDF-Prädikate“ zwar in vielen, aber nicht in allen Fällen möglich; vgl. <http://ta.sandrart.net/de/info/services/lod-rdf/>

innerhalb der Edition vermittelt jedoch zumindest einen Eindruck von dessen Funktionsweise und Ausdrucksmöglichkeiten. Als Beispiel mag an dieser Stelle der Personendatensatz zu Joachim Sandrart dienen.¹³² Dieser enthält verschiedene Personen-Bezüge („Sohn von...“, „Mann von...“, „Schüler von...“ usw.), Kunstwerk-Bezüge („Künstler von...“, „War Stecher von...“, „Ist dargestellt in...“ usw.), Orts-Bezüge („Steht in Zusammenhang mit...“) und Literatur-Bezüge („War Autor von...“, „War Verleger von...“). Das Modell kann also neben eher allgemeinen Zusammenhängen (Verwandtschaft, Autor- und Urheberschaft) auch sehr spezifische, meist in einem kunsthistorischen Kontext relevante Bezüge herstellen, etwa durch Properties respektive Prädikate wie: „War Vermittler(in) des Auftrags für...“ für die Formalisierung des entsprechenden Zusammenhangs zwischen einer Person und einem Kunstwerk (Joachim Sandrart vermittelte [an Artemisia Gentileschi] den Auftrag für das Bild „David mit dem Haupt des Goliath“) oder „Befand sich zu Sandrarts Zeit in...“ und das reziproke „Zu Sandrarts Zeit befand sich hier...“, um den Bezug zwischen einem Kunstwerk und seinem (damaligen) Standort auszudrücken usw. Allerdings wurde die Spezifität nicht bis zum Ausreifungsgrad einer Ontologie ausgeformt. Das Datenmodell von *Sandrart.net* enthält zwar einige allgemeine Kategorien (etwa „Autor“), definiert diese aber nicht formal als Klassen.¹³³

Da das Datenmodell von *Sandrart.net* Aussagen über die Bezüge der Editions-Entitäten untereinander trifft, ist es generell dem Expressivitätsgrad 3 zuzuordnen, auch wenn es auf keine Ontologie zurückgreift und damit nicht den von Stephan Dumont skizzierten Anforderungen für semantisches Kommentieren in digitalen Edition entspricht. In Kombination mit dem Abstraktionsgrad 1 ist *Sandrart.net* also auf Matrixfeld 3 zu verorten. Weitere Projekte, die sich diesem Feld zuordnen lassen, sind die *Bibliothek der Neologie*¹³⁴ und *Der Sturm*.¹³⁵ Allerdings ist einschränkend zu sagen, dass sich zu den Datenmodellen dieser beiden Editionen nur sagen lässt, dass sie auf CIDOC CRM beruhen sollen. Konkretes lässt sich, zumindest auf Grundlage der mir zur Verfügung stehenden Informationen, zum jetzigen Zeitpunkt diesbezüglich nicht ermitteln.

¹³² <http://ta.sandrart.net/de/person/view/4#jumpmark> [09.05.2019]

¹³³ Entsprechend lautete im diesbezüglichen Fragebogen die Antwort auf die Frage, ob bei der Datenmodellierung von *sandrart.net* formale Modelle wie CIDOC CRM verwendet wurden: „Nein, wenngleich unser Datenmodell als CIDOC CRM beeinflusst bezeichnet werden darf.“

¹³⁴ <https://bdn-edition.de> [02.05.2019]

¹³⁵ <https://sturm-edition.de> [02.05.2019]

Ein Abstraktionsgrad höher und damit auf Matrixfeld 6 angesiedelt sind insgesamt drei Projekte, und zwar die *Jahrrechnungen der Stadt Basel 1535-1610 – digital*¹³⁶, die *Urfehdebücher der Stadt Basel – digitale Edition*¹³⁷ und *DEPCHA – Digital Edition Publishing Cooperative for Historical Accounts*.¹³⁸ An dieser Stelle muss angemerkt werden, dass alle diese Editionen unter Mitwirkung des Zentrums für Informationsmodellierung der Universität Graz entstanden sind. Letztlich stellen sie Anwendungsvarianten der dort von Georg Vogeler und anderen entwickelten Bookkeeping Ontology dar (*Baseler Jahrrechnungen*, *DEPCHA*) oder sie bauen doch auf ähnlichen Strukturen auf (*Baseler Urfehdebücher*). Da das Erschließungskonzept der Bookkeeping Ontology schon Gegenstand des vorhergehenden Abschnitts war, soll an dieser Stelle auf eine nähere Beschreibung der *Baseler Jahrrechnungen* und *DEPCHA* zugunsten der *Baseler Urfehdebücher* verzichtet werden.

Quellengrundlage des Projekts sind die 625 Einträge im Baseler Urfehdebuch von 1563 bis 1569. Bei Urfehden handelt es sich um ein verbreitetes Instrument der spätmittelalterlichen und frühneuzeitlichen Rechtspraxis. Ihr Kern bestand darin, dass verurteilte Straftäter durch eine beeidete Erklärung versicherten, am Gericht, das die Strafe verhängte hatte, keine Rache nehmen zu wollen. Durch die schriftliche Fixierung dieser Erklärung sollte die strafverhängende Stadt bzw. ihre Organe und Funktionsträger vor Vergeltung geschützt, das Urteil und die Rechtsprechung als solche legitimiert und der Fall an sich dokumentiert werden. Entsprechend enthalten die Aufzeichnungen neben den meist schematischen Verzichtserklärungen noch eine Reihe kontextualisierender Angaben, die den Fall näher beschreiben. Dazu zählen neben Aussagen über die Straftat selbst sowie das Strafmaß, das nicht selten aus einer Kombination mehrerer Bestrafungen (z.B. Rutenschläge mit anschließendem Stadtverweis oder Geldstrafe mit parallelem Hausarrest usw.) bestand, auch Informationen zu Herkunft, Wohnort, Beruf und Status der Verurteilten und anderer Beteiligten oder Hinweise auf weitere Verfahren enthalten, die mit dem dokumentierten Fall zusammenhängen (z.B. vorangegangene Urfehdeeinträge des Angeklagten). Urfehdebücher enthalten somit zahlreiche Daten, die für die Erforschung der vormodernen Sozial-, Kultur- und Rechtsgeschichte von Interesse sind. Da die Eintragungen ihrem Charakter nach quasi semi-strukturiert sind, eignen sich die Urfehdebücher insbesondere für die vergleichende Analyse großer Einzelfallaggregationen.

¹³⁶ <http://gams.uni-graz.at/context:srbas?mode=projekt> [02.05.2019]

¹³⁷ <https://gams.uni-graz.at/archive/objects/context:ufbas/methods/sdef:Context/get?mode=home> [02.05.2019]

¹³⁸ <http://gams.uni-graz.at/context:depcha> [02.05.2019]

Die Vorteile des Einsatzes von Semantic-Web-Technologien für die Erschließung von Urfehdebüchern und ähnlichen historischen Materialien zu demonstrieren, ist nun erklärtes Ziel der digitalen Edition der Urfehdebücher der Stadt Basel. Dafür wurden erstens unter Verwendung von SKOS Taxonomien erstellt, die alle relevanten Erschließungskategorien definieren und nach Themenbereichen ordnen, z.B. Vergehen, Strafen, Berufe [der beteiligten Personen], Rollen (z.B. „Täter“, „Opfer“, „Notar“, „Mittäter“, „Zeuge“ usw.), aber auch eine Ansammlung ‚gemischter‘ forschungsrelevanter Schlagwörter wie „Armut“, „Blutvogt“, „Erbstreit“, „Halseisen“, „Angaben zu Lohn oder Verdienst“ usw. Zweitens wurde ein auf RDFs basierender Graph erstellt, der als formales Modell zur Abbildung der in den Urfehdebüchern dokumentierten Vorgänge dient.¹³⁹ Herzstück dieses Modells ist die Klasse „Eintrag“, die den Fall selbst repräsentiert und von der ausgehend alle weiteren Properties und Klassen modelliert wurden: das Vergehen, die beteiligten Personen und ihre jeweilige Rolle im Vorgang, die Bestrafung usw.¹⁴⁰ Diese aus ‚Eintrags-Graph‘ und Taxonomien bestehende Modellkombination wird so streng wie möglich vom edierten Text getrennt. Letzterer wird lediglich mit TEI-Markup versehen. Die Verbindung zwischen Markup und Datenmodell wird dann durch das Attribut @ana hergestellt.

Eine dergestalt semantisch angereicherte Edition bietet neben verbesserten Suchfunktionen vor allem den Mehrwert neuer Visualisierungsmöglichkeiten: Die RDF-kodierten Einträge des Urfehdebuchs setzen sich zu einem Gesamtnetzwerk zusammen, aus dem sich komplexe Zusammenhänge ableiten und graphisch darstellen lassen. Diese Art der Informationsvisualisierung, so Christopher Pollin und Georg Vogeler in ihrer Beschreibung des Projekts, eröffnet neue Zugänge für die wissenschaftliche Analyse der Urfehdebücher. Pollin und Vogeler nennen als Beispiel die Fragestellung, ob Frauen für die gleichen Vergehen anders bestraft wurden als Männer.¹⁴¹ An dieser Stelle seien die Möglichkeiten der Graphvisualisierungen etwas ausführlicher demonstriert: Abbildung 5.a stellt den ‚neutralen‘ Graphen für das Vergehen des „Vagierens“ dar. Der große gelbe Knoten repräsentiert das Konzept „female“, das rote Pendant das Konzept „male“. Die Knoten mittlerer Größe repräsentieren verschiedene Straftaten (blau) und Strafen (grün). Die kleinen Knoten stellen die konkreten Einträge im Urfehdebuch dar, wobei jeder Knoten mit dem edierten

¹³⁹ <http://gams.uni-graz.at/archive/objects/o:ufbas.schema/methods/sdef:Ontology/get> [02.05.2019]

¹⁴⁰ Vgl. Pollin et al., *Semantically Enriched Historical Data*, S. 30.

¹⁴¹ Vgl. ebd., S. 31

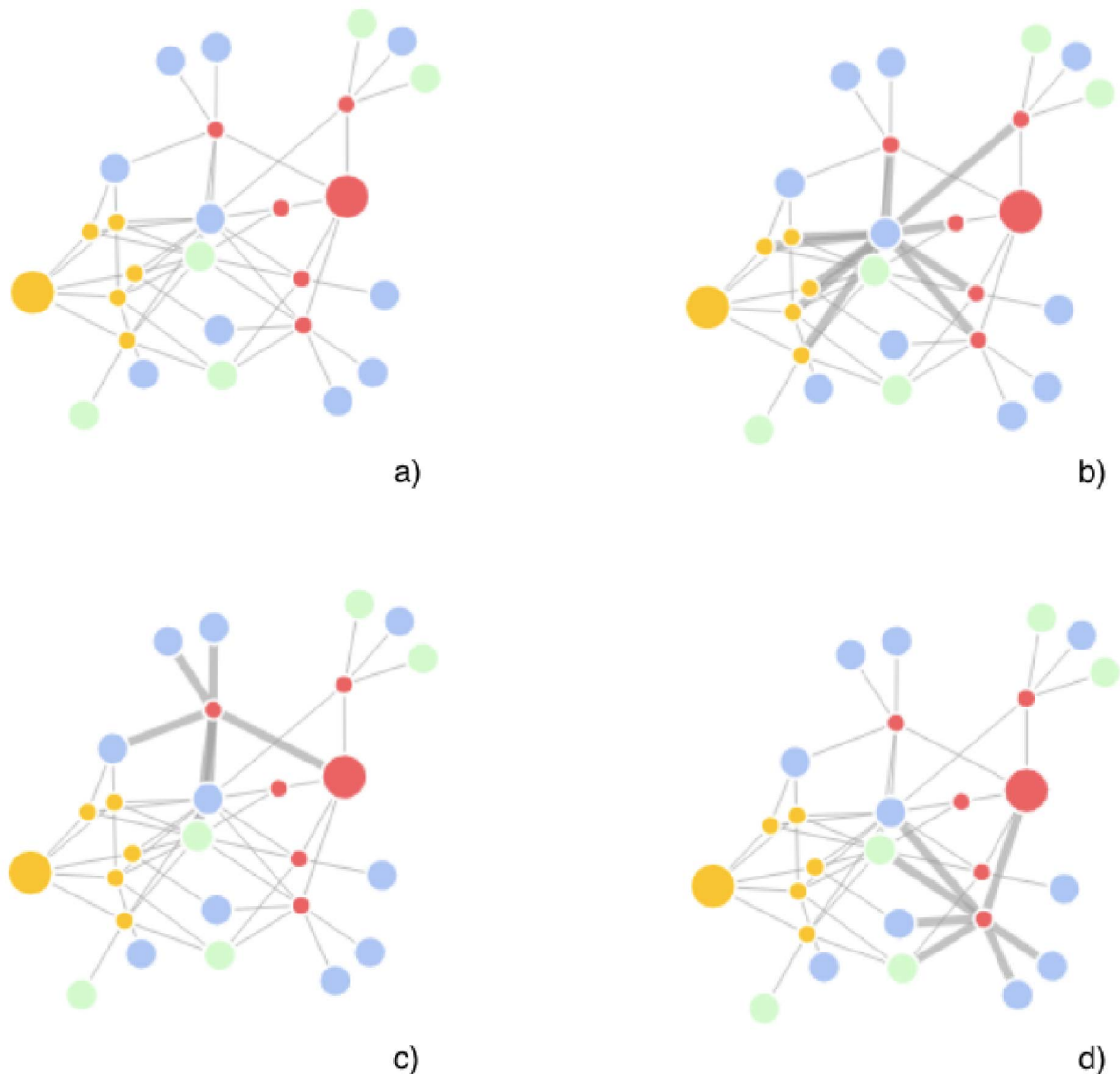


Abb. 5: Graph des Baseler Urfehdebuchs für „Vagieren“ ([https://gams.uni-graz.at/archive/objects/query:ufbas.graph/methods/sdef:Query/get?params=\\$1l%3Chttp://gams.uni-graz.at/ufbas/%23TatVagieren%3E](https://gams.uni-graz.at/archive/objects/query:ufbas.graph/methods/sdef:Query/get?params=$1l%3Chttp://gams.uni-graz.at/ufbas/%23TatVagieren%3E))

Quellentext verbunden ist und per Mausklick aufgerufen werden kann – es ist also jederzeit möglich, von der Grafik- in die Textebene zu wechseln. Die Färbung des Knotens repräsentiert das Geschlecht der angeklagten Person, also rot für männliche und gelb für weibliche Angeklagte. Durch das Anklicken eines beliebigen Knotens wird nun dessen Beziehung zu anderen Knoten des Graphen visualisiert. Abbildung 5.b beispielsweise verdeutlicht den Zusammenhang des „Vagierens“ mit den Einträgen im Urfehdebuch - da dies der Graph für „Vagieren“ ist, ist logischerweise der entsprechende Knoten dieses Vergehens mit allen konkreten Fällen des Graphen verbunden. Darüber hinaus bietet der Graph aber noch eine Reihe weiterer Aufschlüsse, zum Beispiel, dass in einigen Fällen das Vagieren in ‚Tateinheit‘ mit anderen Vergehen (blaue Kreise) steht, etwa Leichtfertigkeit,

Diebstahl und Hehlerei (Abbildung 5.c) oder Ordnungsstörung, Beihilfe und Unzucht (Abbildung 5.d). Die Graphiken zeigen auch, dass diese Fälle von kombinierten Vergehen von männlichen Tätern ausgehen. Eine Untersuchung der ‚weiblichen‘ Einträge würde hingegen zeigen, dass Frauen oft nur mit einem, maximal mit zwei Vergehen in Verbindung gebracht werden.

Der abschließende Blick auf Matrixfeld 9 ist mit einigen Schwierigkeiten verbunden. Zwar würden sich unter Umständen zwei Projekte diesem Feld zuordnen lassen, allerdings nur mit Einschränkungen. Die Gründe sind unterschiedlich: Das erste der beiden Editionsprojekte ist der *Welsche Gast*. Im Rahmen des Projekts werden laut Fragebogen neben Personen und Orten auch Schlagworte erfasst und „wahrscheinlich“ als RDF zur Verfügung gestellt. Darüber hinaus ist ein auf CIDOC CRM basierendes Datenmodell zwar geplant, aber noch nicht umgesetzt. Wie das Datenmodell konkret aussieht und ob es die Kriterien des Matrixfeldes 9 erfüllen würde, kann ich aufgrund der mir zur Verfügung stehenden Informationen¹⁴² allerdings nicht beantworten. Die Einordnung des *Welschen Gastes* auf Matrixfeld geschieht dementsprechend unter großem Vorbehalt. Das Projekt *Stefan George Digital*¹⁴³ hingegen nimmt insofern eine Sonderrolle ein, als es einen eher literaturwissenschaftlichen Zuschnitt besitzt. Herzstück des Projekts ist eine auf Basis von OWL, FRBR und SKOS entwickelte projektspezifische „Schriftontologie“, mit deren Hilfe typographische Phänomene (Typen, Schriftarten, Formen), die bekanntlich bei Stefan George eine zentrale Rolle spielen, formalisiert werden sollen. Das Datenmodell von *Stefan George Digital* greift also offenbar auf eine Ontologie zurück, wendet diese aber nur auf einen sehr speziellen Aspekt an. Um einen solchen Ansatz einzuordnen, ist die von mir verwendete Matrix wenig geeignet. Vor diesem Hintergrund erscheint mir die Einordnung auf Matrixfeld 9 zwar noch als beste, aber keinesfalls unstrittige Lösung.¹⁴⁴

¹⁴² Weder die Projektwebseite noch die Sekundärliteratur können diesbezüglich Aufschlüsse liefern.

¹⁴³ Die Edition ist ein in Arbeit befindliches und daher bisher unveröffentlichtes Promotionsprojekt; vgl. bis auf Weiteres Frederike Neuber: Stefan George Digital. Poster at the 3rd DHd conference in Leipzig, 2016, S. 357–359. <http://www.dhd2016.de/abstracts/posters-033.html#index.xml-body.1>.

¹⁴⁴ Ebenfalls denkbar wäre z.B. eine Einordnung in die Felder 3 oder 8.

5. Fazit

Auch wenn der seit Jahren antizipierte Durchbruch des Semantic Web nach wie vor auf sich warten lässt, haben Semantic-Web-fähige Datensammlungen und Strukturen bis dato einen beträchtlichen Verbreitungsgrad erreicht. Dies gilt, bis auf wenige Ausnahmen, auch und gerade für den Kulturerbebereich. Das Editions Wesen zählte, obwohl die Anwendungsmöglichkeiten vielfältig und der avisierte Nutzen beträchtlich sind, bis vor Kurzem zu diesen Ausnahmen. Und auch nach jetzigem Stand kann beileibe nicht davon gesprochen werden, dass elaborierte Semantic-Web-Anwendungen zum Standardrepertoire digitaler Editionen gehören. Entsprechende Szenarien sind in der Fachliteratur zwar unter verschiedenen Vorzeichen erörtert worden, in der Praxis spielen sie aber noch eine untergeordnete Rolle: Zwar setzten von den mir untersuchten 52 Editionen 47 Normdaten ein, allerdings beschränkt sich der semantische Mehrwert beim Großteil der Projekte auf die eindeutige Referenzierung von Entitäten.

Andererseits kann angesichts von 17 Projekten, die in der einen oder anderen Form an das Semantic Web anknüpfen, auch nicht (mehr) von einem Nischenphänomen die Rede sein. Freilich zeichnet sich hier noch kein ‚Königsweg‘ oder Universal-Modell zur Integration digitaler Editionen in das Semantic Web ab. Von einer gewissen praktischen Relevanz ist aber offensichtlich der bei *Burckhardtsource*, *Theodor Fontane: Notizbücher* und der *Carl-Maria-von-Weber-Gesamtausgabe* bereits umgesetzte und für einige weitere Projekte in Planung befindliche Ansatz, über Normdatensätze einen ‚Anschluss‘ an Linked-Open-Data-fähige Datenräume herzustellen. Zur Attraktivität scheint neben der Möglichkeit der projektübergreifenden Identifizierung ein gewisses ‚ready-to-use‘-Moment beizutragen, das sich aus den semantischen Vernetzungen der Normdaten ergibt. Ganz im Sinne des Semantic Web lassen sich durch die Zusammenführung ursprünglich verteilter Informationen implizite Bezüge explizieren, etwa wenn in einem Brief erwähnte Personen nicht nur gekennzeichnet, sondern auch nach ihrem – im Brief nicht genannten – Geburtsort geordnet werden können.

Einen Schritt weiter gehen in dieser Beziehung Projekte wie *Sandart.net*: Projektspezifische Datenmodelle erlauben neue, komplexere Analysezugänge zum edierten Material, etwa durch erweiterte Suchfunktionen und Visualisierungsmöglichkeiten. Und wie die *Baseler Urfehdebücher* oder die *Baseler Jahrrechnungen* demonstrieren, lässt sich dieses Konzept nicht nur auf konkrete, sondern auch auf abstrakte Entitäten ausdehnen. Vorerst allerdings ist der Anwendungsbereich solcher Strukturen, die aus der Kombination eines domänenspezifischen Datenmodells wie der wirtschaftshistorischen Bookkeeping Ontolo-

gie oder dem ‚kriminalhistroschen‘ Urfehde-Graphen mit diversen thematischen Taxonomien besteht, vorerst auf die Erschließung hoch strukturierten Schriftguts beschränkt.

Damit ist freilich die Frage, welchen Mehrwert Semantic-Web-Technologien für digitale Editionen bieten können, nicht abschließend beantwortet. Denn zum ersten bleibt es abzuwarten, ob es gelingt, die elaboriertesten Ansätze für die Nutzung von Semantic-Web-Technologien – z.B. die Erschließung abstrakter Konzepte in unstrukturierten Textmaterialien wie Briefen – auch in der Praxis umzusetzen. Dieser Frage, die gewissermaßen die Dimension der *Tiefe* der Erschließung berührt, ist mit einem zweiten Problemkomplex verbunden, der die *Breite* der Erschließung betrifft: Der Mehrwert von Semantic-Web-Technologien hängt unmittelbar vom Grad ihrer Verbreitung ab, ihr volles Nutzungspotenzial entfalten sie erst ab einem kritischen Punkt. Dass dieser Punkt im digitalen Editionswesen erreicht wäre, kann freilich nicht behauptet werden: Normdaten wie die Personendatensätze der GND oder die Ortsdatensätze der GeoNames sind der einzige flächendeckend verbreitete Standard, der – zumindest im Prinzip – auf Semantic-Web-Strukturen aufbaut. Eine solche flächendeckende Verbreitung weiterer Standards wäre aber die Voraussetzung dafür, dass Wissen zwischen den verschiedenen ‚Datenräumen‘ – also zwischen den Editionen selbst sowie zwischen den Editionen einerseits und anderen ‚Datenräumen‘ andererseits – vernetzt werden und das Semantic Web sein volles Potenzial entfalten könnte.

Ob Semantic-Web-Technologien in dieser Tiefe und Breite jemals Anwendung finden werden, kann hier natürlich nicht beantwortet werden, da dies nicht zuletzt von den Erwartungen der Nutzer an die Leistung einer digitalen Edition abhängt. Ob vor diesem Hintergrund die Fachcommunity ein Semantic-Web-basiertes Editions-Szenario überhaupt vielversprechend findet oder ob sie in den hier skizzierten Ansätzen keinen entscheidenden, den Einsatz relativ komplexer Technologien rechtfertigenden Mehrwert zu erblicken vermag, ist aber eine Frage, der nachzugehen außerhalb der Möglichkeiten der vorliegenden Arbeit liegt.

Verzeichnis der verwendeten Literatur

Albers, Laura; Große, Peggy; Wagner, Sarah: Semantische Datenmodellierung mit CIDOC CRM – Drei Fallbeispiele, in: Maximilian Eibl, Martin Gaedke (Hgg.), Informatik 2017. Tagung 2017 in Chemnitz, Band 2 [= Jahrestagung der Gesellschaft für Informatik 47], Bonn 2017, S. 1209-1220. DOI: <https://doi.org/10.11588/artdok.00005745>.

Aliverti, Christian; Fabian, Claudia; Kailus, Angela Kailus: RDA und Kultureinrichtungen, in: Zeitschrift für Bibliothekswesen und Bibliographie, 62 2015, Heft 6, S. 329-338. <http://dx.doi.org/10.3196/186429501562643>.

Augustin, Christian; Morgenstern, Ulf; Riechert, Thomas: Der Leipziger Professorenkatalog. Ein Anwendungsbeispiel für kollaboratives Strukturieren von Daten und zeitnahe Publizieren von Ergebnissen basierend auf Technologien des Semantic Web und einer agilen Methode des Wissensmanagements, in: Schomburg, Silke; Leggewie, Claus; Lobin, Henning; Puschmann, Cornelius (Hgg.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland, Köln 2011, S. 41–44.

Becker, Irmgard Christa: Archivische Erschließung und RDA. Stellungnahme der Archivschule Marburg zur Nutzung der RDA (Resource Description and Access) in Archiven. http://www.dnb.de/SharedDocs/Downloads/DE/DNB/standardisierung/rdaKultur2013Becker.pdf?__blob=publicationFile.

Berners-Lee, Tim; Handler, James; Lassila, Ora: The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, in: Scientific American 284 (2001), Issue 5, S. 34–43; dt. als: Mein Computer versteht mich. In: Spektrum der Wissenschaft, August 2001, S. 42–49.

Boltzendahl, Sabine: Ontologien in digitalen Bibliotheken unter dem Schwerpunkt Inhaltsserschließung und Recherche [= Berliner Handreichungen zur Bibliothekswissenschaft 111]. Berlin 2013. DOI: 10.18452/18353.

Chen, Esther: Linked Open VD 17 – von METS/MODS zum Europeana Data Model. Überlegungen zum technischen Migrationspfad und zum funktionalen Mehrwert semantischer Nutzungsszenarien [= Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft 327]. Berlin 2012, S. 17. DOI: 10.18452/2069.

Ciotti, Fabio; Tomasi, Francesca: Formal Ontologies, Linked Data, and TEI Semantics, in: Journal of the Text Encoding Initiative 9 (2016/17). DOI: 10.4000/jtei.1480.

Ciotti, Fabio; Lana, Maurizio; Francesca Tomasi: TEI, Ontologies, Linked Open Data. Geolat and Beyond, in: Journal of the Text Encoding Initiative 8 (2014 /15). DOI: 10.4000/jtei.1365.

Ciula, Arianna; Viera, Jose Miguel: Implementing an RDF/OWL Ontology on Henry the III Fine Rolls, in: OWL. Experiences and Directions Third International Workshop 2007. <http://www.cidoc-crm.org/Irmoo/Resources/implementing-an-rdfowl-ontology-on-henry-the-iii-fine-rolls>.

Dies.; Spence, Paul: Expressing complex associations in medieval historical documents – The Henry III Fine Rolls Project, in: Literary and Linguistic Computing. The journal of digital scholarship in the humanities 23 (2008), No. 3, S. 311–325. <https://doi.org/10.1093/lc/fqn018>.

DeRose, Steven J.; Durand, David G.; Mylonas, Elli; Renear, Allen H.: What is text, really? In: Journal of Computing in Higher Education 1,2 (1990), S. 3–26. <https://doi.org/10.1007/BF02941632>.

Doerr, Martin; Gradmann, Stefan; Hennicke, Steffen; Isaac, Antoine; Meghini, Carlo; Sompel, Herbert van de: The Europeana Data Model (EDM). IFLA 2010, World Library and Information Congress: 76th IFLA General Conference and Assembly, 2010, Gothenburg. URL: <<https://www.ifla.org/past-wlic/2010/149-doerr-en.pdf>>.

Donato, Francesca di; Morbidoni, Christian; Fonda, Simone; Piccoli, Alessio; Gratis, Marco; Nucci, Michele: Semantic Annotation with Pundit. A case study and a practical demonstration, in: Proceedings of DH-Case 2013, Florence, 2013. DOI: 10.1145/2517978.2517995.

Dumont, Stefan: Briefe kommentieren im Semantic Web – ein Konzept, in: DARIAH-DE Working Papers Nr. 33. Göttingen 2019. urn:nbn:de:gbv:7-dariah-2019-5-8.

Efer, Thomas: Graphdatenbanken für die textorientierten e-Humanities. Leipzig 2017. urn:nbn:de:bsz:15-qucosa-219122.

Ders.: Introducing NoXML for the Digital Humanities, in: Proceedings of INFORMATIK 2017, Lecture Notes in Informatics (LNI), hg. von der Gesellschaft für Informatik, 2017, S. 1161–1666. DOI:10.18420/in2017_117.

Eide, Øyvind: Ontologies, Data Modeling, and TEI, in: Journal of the Text Encoding Initiative 8 (2014/15). DOI: 10.4000/jtei.1191.

Gruber, Thomas R.: A Translation Approach to Portable Ontology Specifications, in: Knowledge Acquisition 5 (1993), Issue 2, S. 199–220.

Hauke, Petra; Grünewald, Jana; Kader, Ben; Kaufmann, AndreaM Kindling, Max; Voss, Jakob: Das Imageproblem der deutschen Bibliothekswissenschaft. Vortragsfolien für den 94. Deutschen Bibliothekartag 2005. <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/90>.

Hauser, Julia; Heuvelmann, Reinhold; Rühle, Stefanie; Voß, Jakob: Interoperable Metadaten im Giant Global Graph, in: Dialog mit Bibliotheken 2014, Heft 2, S. 41–46

Heßbrüggen-Walter, Stefan: Tatsachen im semantischen Web: Nanopublikationen in den digitalen Geisteswissenschaften?, in: Peter Haber, Eva Pfanzelter (Hgg.), Historyblogosphere. Bloggen in den Geschichtswissenschaften, München 2013, S. 149-160. DOI: 10.1524/9783486755732.149

Ders.: What People Said: The Theoretical Foundations of a Minimal Doxographical Ontology and Its Use in the History of Philosophy, in: Constanze Baum, Thomas Stäcker (Hgg.), Grenzen und Möglichkeiten der Digital Humanities [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 1]. Wolfenbüttel 2015. DOI: 10.17175/sb001_001.

Hennicke, Steffen: Linked Data und semantische Suchfunktionalität in Europeana, in: Mitteilungen der VÖB 66 (2013), Heft 1, S. 20–34

Hohmann, Georg: Die Anwendung von Ontologien zur Wissensrepräsentation und -kommunikation im Bereich des kulturellen Erbes, in: Schomburg, Silke; Leggewie, Claus; Lobin, Henning; Puschmann, Cornelius (Hgg.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland, Köln 2011, S. 33–39.

Iglesia, Martin de la; Moretto, Nicolas; Brodhun, Maximilian: Metadaten, LOD und der Mehrwert standardisierter und vernetzter Daten, in: Heike Neuroth, Andrea Rapp, Sibylle Söring (Hgg.), TextGrid: Von der Community – für die Community, Glückstadt 2015, S. 91–104.

Jordanous, Anna; Stanley, Alan; Tupman, Charlotte: Contemporary transformation of ancient documents for recording and retrieving maximum information: when one form of markup is not enough, in: *Proceedings of Balisage: The Markup Conference 2012*. Balisage Series on Markup Technologies 8 (2012). <https://doi.org/10.4242/BalisageVol8.Jordanous01>.

Kamzelak, Roland S.: Chancen und Grenzen von Normdaten, FRBR und RDF, in: „Ei, dem alten Herrn zoll’ ich Achtung gern“, Festschrift für Joachim Veit zum 60. Geburtstag, München 2016, S. 423-435. DOI: 10.25366/2018.29.

Kuczera, Andreas: Digital Editions beyond XML – Graph-based Digital Editions, in: Histoinformatics 2016 - The 3rd Histoinformatics Workshop. Proceedings of the 3rd Histoinfor-

matics Workshop on Computational History (Histoinformatics 2016), hrsg. von Marten Düring, Adam Jatowt, Johannes Preiser-Kappeller, co-located with Digital Humanities 2016 conference (DH 2016), Krakow, 2016, S. 37-46. http://ceur-ws.org/Vol-1632/paper_5.pdf.

Ders.: Graphentechnologien in den Digitalen Geisteswissenschaften, in: ABI Technik 37 (2017), Heft 3, S. 179–196. DOI: 10.1515/abitech-2017-0042.

Ders.; Wübbena, Torsten; Kollatz, Thomas (Hgg.): Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten [= Zeitschrift für Ideengeschichte, Sonderband 4], Wolfenbüttel 2019. DOI: 10.17175/sb004.

Kummer, Robert: Semantic Technologies for Manuscript Descriptions — Concepts and Visions, in: Franz Fischer, Christiane Fritze, Georg Vogeler (Hgg.), Kodikologie und Paläographie im digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2 [= Schriften des Instituts für Dokumentologie und Editorik 3], Norderstedt 2011, S. 133–154. <http://kups.ub.uni-koeln.de/id/eprint/4347>.

Münnich, Stefan: Ontologien als semantische Zündstufe für die digitale Musikwissenschaft? Eine Bestandsaufnahme, in: Bibliothek, Forschung und Praxis 42 (2018), Heft 2, S. 184–193. DOI: 10.18452/18946.

Neuber, Frederike: Stefan George Digital. Poster at the 3rd DHd conference in Leipzig, 2016, S. 357–359. <http://www.dhd2016.de/abstracts/posters-033.html#index.xml-body.1>.

Ore, Christian-Emil; Eide, Øyvind: TEI and cultural heritage ontologies: Exchange of information?, in: Literary and Linguistic Computing, Vol. 24, No. 2, 2009, S. 161-172. <https://doi.org/10.1093/lc/fqp010>.

Pohl, Adrian; Danowski, Patrick: Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick, in: dies. (Hgg.), (Open) Linked Data in Bibliotheken, Berlin, Boston 2013, S. 1–44. DOI: 10.1515/9783110278736.1.

Pollin, Christopher; Vogeler, Georg: Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the "Urfehdebücher der Stadt Basel", in: Proceedings of the Second Workshop on Humanities in the Semantic Web, co-located with 16th International Semantic Web Conference (DH 2017), Wien, 2017, S. 27–32. <http://ceur-ws.org/Vol-2014/paper-03.pdf>.

Rehbein, Malte: Ontologien, in: Fotis Jannidis, Hubertus Kohle, Malte Rehbein (Hgg.), Digital Humanities. Eine Einführung, Stuttgart 2017, S. 162–176.

Sahle, Patrick: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels [= Schriften des Instituts für Dokumentologie und Editorik 7-9], 3 Bände, Norderstedt 2013.

Schaßan, Torsten: Digitale Quellen – Datei- und Datenformate, in: Laura Busse, Wilfried Enderle, Rüdiger Hols, Gregor Horstkemper, Thomas Meyer, Jens Prellwitz, Annette Schuhmann (Hgg.), Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften [= Historisches Forum 19], Berlin 2016, A.6 – 21. <http://www.clio-online.de/guides/arbeitsformen-und-techniken/digitale-quellen-dateiformate/2016>.

Schrade, Torsten: Geisteswissenschaftliche Fachdatenrepositorien im Semantic Web. Modellierung, Vernetzung, Visualisierung. Conference paper and presentation at the 3rd DHd conference in Leipzig, 2016, S. 232–239. <http://www.dhd2016.de/abstracts/vorträge-054.html>.

Šimek, Jakob: Der Nutzen graphbasierter Datenmodelle für digitale Editionen. Unveröffentlichte Masterarbeit. Berlin 2018.

Stadler, Peter: Normdateien in der Edition, in: Editio 26 (2012), Heft 1, S. 174–183. DOI: <https://doi.org/10.1515/editio-2012-0013>.

Stein, Christian: Linked Open Data – Wie das Web zur Semantik kam, in; Bibliothek, Forschung und Praxis 38 (2014), Heft 3, S. 1–9. https://www.interdisciplinary-laboratory.hu-berlin.de/de/content/linked-open-data-wie-das-web-zur-semantik-kam/?content_type=mitglied.

Strobel, Jochen: Digitale Briefedition und semantische Erschließung. Von den Briefen der Jenaer Romantikergeneration zur Edition der Korrespondenz August Wilhelm Schlegels, in: editio 28 (2014), S. 151–174.

Thaller, Manfred: What is a text within the Digital Humanities, or some of them, at least? In: digital humanities 2012. Hamburg 2012. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/beyond-embedded-markup/>.

Tummarello, Giovanni; Morbidoni, Christian; Kepler, Fabio N.; Piazza, Francesco; Puliti, P.: A novel Textual Encoding paradigm based on Semantic Web tools and semantics, in: Proceedings of the 5th Edition of the International Conference on Language Resources and Evaluation, Paris 2006, S. 247–252. http://www.lrec-conf.org/proceedings/lrec2006/pdf/225_pdf.pdf.

Vetter, Johannes: Semantik Web im Überblick. https://www.researchgate.net/publication/267845680_Semantic_Web_im_Uberblick.

Vlachidis, Andreas; Bikakis, Antonis; Kyriaki-Manessi, Daphne; Triantafyllou, Ioannis; Padfield, Joseph; Kontiza, Kalliopi: Semantic Representation and Enrichment of Cultural Heritage Information for Fostering Reinterpretation and Reflection on the European History, in: Marinos Ioannides (Hg.), *Digital Cultural Heritage*, Cham 2018, S. 91–103.

Vogeler, Georg: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?, in: Constanze Baum, Thomas Stäcker (Hgg.), *Grenzen und Möglichkeiten der Digital Humanities* [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 1]. Wolfenbüttel 2015. DOI: 10.17175/sb001_007.

Ders.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, in: Jürgen Sarnowsky (Hg.), *Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters*, Göttingen 2016, S. 13-42. DOI: 10.14220/9783737006774.13.

Ders.: The 'assertive edition' : On the consequences of digital methods in scholarly editing for historians, in: *International Journal of Digital Humanities* 1 (2018). <http://dx.doi.org/10.17613/M6JS9H76P>.

Wang, Yongming; Yang, Sharon Q.: Linked Data Technologies and What Libraries Have Accomplished So Far, in: *International Journal of Librarianship*, 3 (2018), Issue 1, S. 3–20. DOI: <https://doi.org/10.23974/ijol.2018.vol3.1.62>.

Wettlaufer, Jörg; Johnson, Christopher; Scholz, Martin; Fichtner, Mark; Thotempudi, Sree Ganesh: Semantic Blumenbach. Exploration of Text–Object Relationships with Semantic Web Technology in the History of Science, in: *Digital Scholarship in the Humanities* 30 (2015), Issue suppl. 1, S. 187–198. <https://doi.org/10.1093/lc/fqv047>.

Ders.: Der nächste Schritt? Semantic Web und digitale Editionen, in: Roland S. Kamzelak, Timo Steyer (Hgg.), *Digitale Metamorphose: Digital Humanities und Editionswissenschaft* [= Zeitschrift für digitale Geisteswissenschaften, Sonderband 2]. Wolfenbüttel 2018. DOI: 10.17175/sb03.

Verzeichnis der verwendeten Internetseiten

Akademie der Wissenschaften und der Literatur / Marjam Trautmann und Torsten Schrade: Der Sturm. Digitale Quellenedition zur Geschichte der internationalen Avantgarde [<https://sturm-edition.de>]

Berlin-Brandenburgische Akademie der Wissenschaften: correspSearch. Briefeditionen durchsuchen und vernetzen [<https://correspsearch.net/index.xql?l=de>]

Berners-Lee et al.: Uniform Resource Identifier (URI): Generic Syntax [<https://tools.ietf.org/html/rfc3986>]

BIB e.V.: Konferenzveröffentlichungen (Vortragsfolien) zu den Deutschen Bibliothekartagen / "Semantic Web" [https://opus4.kobv.de/opus4-bib-info/solrsearch/index/search/start/0/rows/20/sortfield/score/sortorder/desc/searchtype/simple/query/semantic+web/facet-Number_year/all/doctypesfq/Konferenzfolien]

BIB e.V.: Konferenzveröffentlichungen (Vortragsfolien) zu den Deutschen Bibliothekartagen / "Linked Open Data" [<https://opus4.kobv.de/opus4-bib-info/solrsearch/index/search/start/0/rows/20/sortfield/score/sortorder/desc/searchtype/simple/query/Linked+Open+Data/doctypesfq/Konferenzfolien>]

burckhardtsource project / Maurizio Ghelardi: Burckhardtsource / Home [<https://www.burckhardtsource.org>]

Census of Antique Works of Art and Architecture Known in the Renaissance (Hu Berlin): Startseite [<http://www.census.de/census>]

Census of Antique Works of Art and Architecture Known in the Renaissance (Hu Berlin): Kulturerbe als Linked Open Data - Pilotprojekt für den Census [<http://www.census.de/census/nachrichten/kulturerbe-als-linked-open-data-pilotprojekt-fuerden-Census>]

Deutsche Forschungsgemeinschaft / Gruppe Informationsmanagement: GEPRIS - Geförderte Projekte der DFG [<https://gepris.dfg.de/geprisOCTOPUS;jsessionid=813C8093B23C03FBD5->

Deutsche Nationalbibliothek: GND Ontology / Description Document 2019-10-15 [<https://d-nb.info/standards/elementset/gnd>]

Deutsche Schillergesellschaft e.V. / Deutsches Literaturarchiv Marbach: Digital Humanities – Projekte [<https://www.dla-marbach.de/digital-humanities/projekte/>]

Digital Humanities Lab, University of Basel: NIE-INE / Nationale Infrastruktur für Editionen [<https://www.nie-ine.ch>]

Early Modern Thought Online / Stefan Heßbrüggen-Walter: EMTO Nanopub [http://emto-nanopub.referata.com/wiki/EMTO_Nanopub]

Europeana Foundation: Europeana [<https://www.europeana.eu/portal/de>]

Europeana Foundation: Europeana data model [<https://www.europeana.eu/portal/de>]
FA77901355C7C6]

Georg-August-Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek Göttingen: Bibliothek der Neologie [<https://bdn-edition.de>]

GVogeler: bookkeeping / Ontology for the description of the content of historical accounting documents [<https://github.com/GVogeler/bookkeeping>]

Herzog August Bibliothek Wolfenbüttel: Sandart, Joachim von / Basis-Daten [<http://ta.-sandart.net/de/person/view/4#jumpmark>]

Herzog August Bibliothek Wolfenbüttel: Sandart.net: Eine netzbasierte Forschungsplattform zur Kunst- und Kulturgeschichte des 17. Jahrhunderts [www.sandart.net/de/]

Herzog August Bibliothek: Liste aller digitalen Editionen [<http://www.hab.de/de/home/bibliothek/digitale-bibliothek-wdb/digitale-editionen/liste-allerdigitalen-editionen.html>]

Historische Kommission bei der Bayerischen Akademie der Wissenschaften: Deutsche Biografie [<https://www.deutsche-biographie.de/>]

ICOM/CIDOC Documentation Standards Group / CIDOC CRM Special Interest Group: Definition of the CIDOC Conceptual Reference Model / Version 5.0.4 [http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf]

IFLA: FRBRoo [<https://www.ifla.org/node/10171>]

Insight Centre for Data Analytics: The Linked Open Data Cloud [<https://lod-cloud.net/>]
International Organization for Standardization (ISO): ISO 21127:2014 / Information and documentation — A reference ontology for the interchange of cultural heritage information [<https://www.iso.org/standard/57832.html>]

Journal of the TEI / Johan Walsh: Journal of the Text Encoding Initiative [<https://journal.tei-c.org/index.php/journal/index>]

Österreichische Akademie der Wissenschaften: Catalogue Digital Editions [<https://dig-ed-cat.acdh.oeaw.ac.at/>]

Patrick Sahle: (A catalog of) Digital Scholarly Editions [<http://www.digitale-edition.de/>]
Stiftung Preußischer Kulturbesitz / Deutsche Digitale Bibliothek: Deutsche Digitale Bibliothek [<https://www.deutsche-digitale-bibliothek.de>]

Susanna Burghartz, Departement Geschichte, Universität Basel / Georg Vogeler, Zentrum für Informationsmodellierung - Austrian Centre for Digital Humanities, Karl-Franzens-Universität Graz: Urfehdebücher der Stadt Basel – digitale Edition [<https://gams.uni-graz.at/archive/objects/context:ufbas/methods/sdef:Context/get?mode=home>]

Text Encoding Initiative Consortium: TEI. TEI Lite [<http://www.tei-c.org/guidelines/customization/lite/>]

Text Encoding Initiative Consortium: TEI. Text Encoding Initiative [<http://www.tei-c.org>]

Union der deutschen Akademien der Wissenschaften e.V.: Forschungsprojekte im Akademienprogramm [https://www.akademienunion.de/forschungsprojekte/?no_cache=1]

University of Graz / Georg Vogeler: DEPCHA [<http://gams.uni-graz.at/context:depcha>]

Wikipedia: Linked Open Data [https://de.wikipedia.org/wiki/Linked_Open_Data]

World Wide Web Consortium / W3C JSON-LD Working Group: JSON for Linking Data [<https://json-ld.org>]

World Wide Web Consortium: Notation3 (N3): A readable RDF syntax [<https://www.w3.org/TeamSubmission/n3/>]

World Wide Web Consortium: OWL Ontology [https://www.w3.org/standards/techs/owl#w3c_all]

World Wide Web Consortium: RDF 1.1 Turtle / Terse RDF Triple Language / W3C Recommendation 25 February 2014 [<https://www.w3.org/TR/turtle/>]

World Wide Web Consortium: RDF 1.1 XML Syntax / W3C Recommendation 25 February 2014 [<https://www.w3.org/TR/rdf-syntax-grammar/>]

World Wide Web Consortium: RDF Schema 1.1 / W3C Recommendation 25 February 2014 [<https://www.w3.org/TR/rdf-schema/>]

World Wide Web Consortium: Resource Description Framework (RDF) [<https://www.w3.org/RDF/>]

World Wide Web Consortium: SKOS Simple Knowledge Organization System Reference / W3C Recommendation 18 August 2009 [<https://www.w3.org/TR/skos-reference/>]

World Wide Web Consortium: SPARQL Query Language for RDF / W3C Recommendation 15 January 2008 [<https://www.w3.org/TR/skos-reference/>]

World Wide Web Consortium: SPARQL Query Language for RDF / W3C Recommendation 15 January 2008 [<https://www.w3.org/TR/rdf-sparql-query/>]

XBRL Deutschland e.V.: XBRL [<https://de.xbrl.org>]

Zentrum für Informationsmodellierung - Austrian Centre for Digital Humanities: Jahrrechnungen der Stadt Basel 1535 bis 1610 – digital / Beta-Version [<http://gams.uni-graz.at/context:srbas?mode=projekt>]

Anhang

Tabelle: Korpus der untersuchten Editionsprojekte (Angaben aus den Fragebögen in normaler Schrift; aus Webseiten und Sekundärliteratur gezogene Angaben kursiv; nicht ermittelbare Angaben mit „XXX“ gekennzeichnet)

Kurzname	Projektname	Projekt anfang	Projek- tende	Markup	Normdaten	Entitäten (1/Zusatz/3)	Verwendung von RDF?	Verwendung formaler Modelle?	Matrixfeld
Abacus	ABaC:us - Austrian Baroque Circus	2015	n.n.	TEI	nein	Personen, Orte	nein	nein	1
ARCHITRAVE	Kunst und Architektur in Paris und Versailles im Spiegel deutscher Reiseberichte des Barock	2017	2020	TEI	Getty Union List of Artist Names, VIAF, Getty Thesaurus of Geographic Names	Personen, Orte, Werke	nein	nein	1
	Die Augsburger Baumeisterbücher: digitale Edition, Kommentar und Präsentation der mittelalterlichen Stadtrechnungen von 1320 bis 1466	2014	n.n.	TEI	nein	Namen, Orte, Körperschaften, Schlagworte	nein	nein	4
Austrian precursor	The numismatic networks of Eckhel's Austrian precursors	2017	2020	TEI	GND, Geonames	Personen, Orte, Werke	Geplant	Nein	2 (geplant)
Bahr & Schnitzler	Hermann Bahr - Arthur Schnitzler: Briefwechsel, Aufzeichnungen, Dokumente (1891-1931)	2016	2018	TEI	nein	Personen, Werke, Orte, Organisationen	nein	nein	1
Basel Jahresrechnungen	Jahresrechnungen der Stadt Basel 1535-1610 – digital	2014	2015	TEI	nein, aber URIs	Rechnungsentitäten: Buchungskategorien, Buchungseinträge, Geldbeträge, Summen	RDF	nein, aber Bookkeeping Ontology	6
Basel Urfehdebücher	Urfehdebücher der Stadt Basel - digitale Edition	2012	2016	TEI	nein, aber URIs	Personen, Orte, „Schlagworte“ als Repräsentation der im Urfehdebuch dokumentierten Rechtsfälle, also eigentlich: Täter, Opfer, Taten, Urteile, Berufe, Herkunftsorte	RDF	nein, aber Taxonomien	6
Benigna von Reuß-Ebersdorf	Digitale Edition der Briefe Erdmuthes Benignas von Reuß-Ebersdorf (1670-1732)	2015	TEI	GND	Personen, Orte, Organisationen, Schlagwörter („zentrale Themenkomplexe“)	nein	nein	nein	7
Benn & Zenzes	Kommentierte Edition des Briefwechsels zwischen Gottfried Benn und Gertrud Zenzes	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Berlin	Briefe und Texte aus dem intellektuellen Berlin um 1800	2010	2015	TEI	GND	Personen, Werke, Orte, Organisationen	nein	nein	1
Blumenbach	Blumenbach – online [& Semantic Blumenbach]	2010	2024	TEI	GND, Getty Thesaurus of Geographic Names	Personen, Orte, Werke, Artefakte	[RDF]	[CIDOC CRM]	1[*]
Borchward	Borchwards Reise	2014	2015	„Editionszeichen“	GND	Personen	RDFa; aber kein SPARQL-Endpunkt	Nein	1
Burckhardt	Burckhardtsource	2010	2015	Ja	GND	Personen, Orte, Institutionen, (bibl.) Werke, Daten, Kunstwerke	RDF (via Pundit)	Nein	2
Christian II.	Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg (1599-1656)	2013	2025	TEI	GND	Personen, Orte, Körperschaften, Schlagwörter, Bibelstellen	nein	Nein	7

DEPCHA	Digital Edition Publishing Cooperative for Historical Accounts
Dumont & Lindemann	Louise Dumont an Gustav Lindemann – Eine digitale Briefedition
Eckhel	Joseph Eckhel and his numismatic Network
Escher	Alfred Escher-Briefedition
Exilbriefnetz	Vernetzte Korrespondenzen: Exilbriefnetz
Faulhaber	Kritische Online-Edition der Tagebücher Michael Kardinal von Faulhabers
Fontane	Theodor Fontane: Notizbücher. Digitale genetisch-kritische und kommentierte Edition
Friedenswahrung	Religiöse Friedenswahrung und Friedensstiftung in Europa (1500- 1800):Digitale Quellenedition
Friedrich der Weise	frühneuzeitlicher Religionsfrieden Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johanns des Beständigen 1513 bis 1532. Reformation im Kontext frühneuzeitlicher Staatswerdung
Fürstinnenkorresponden z	Frühneuzeitliche Fürstinnenkorrespondenz im mitteldeutschen Raum
Gerken	Der erste 'Baedeker' von Berlin. Die Stadtbeschreibung von Johan Heinrich Gerken 1714-1717
George	Stefan George Digital
Goethe & Riemer	Johann Wolfgang Goethe. Briefwechsel mit Friedrich Wilhelm Riemer (Hybridedition)
Haeckel	Ernst Haeckel (1834-1919): Briefedition
Hainhofer	Kommentierte digitale Edition der Reise- und Sammlungsbeschreibungen Philipp Hainhofers (1578–1647)
Haller	Online-Edition der Rezensionen und Briefe Albrecht von Hallers

2018	<i>n.n.</i>	TEI	nein (aber URIs)	Rechnungsentitäten: Buchungskategorien, Buchungseinträge, Geldbeträge, Summen	RDF	nein, aber bookkeeping ontology	6
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2012	2015	TEI	GND, GeoNames	Personen, Orte, Werke	geplant	nein	2 (geplant)
2008	<i>n.n.</i>	<i>angelehnt an TEI</i>	GND	<i>Personen, Orte, Körperschaften, Schlagwörter, Bibelstellen</i>	<i>nein</i>	<i>Nein</i>	7
2013	<i>n.n.</i>	TEI	GND, GeoNames	Personen, Orte, Werke, Periodika, Körperschaften, Schlagwörter	nein	nein	7
2014	2015	eingenes XML	GND/VI/AF	Personen, Dokumente	nein	nein	1
2015?		TEI	GND	<i>Personen, Ereignisse, Institutionen / Körperschaften, Werke, Orte</i>	<i>RDF (via TextGrid)</i>	<i>nein</i>	2
2013	2019	TEI	GND, GeoNames	Personen, Orte, Werke	nein	nein	1
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2010	2013	nein	nein	nein	nein	nein	---
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2015	2020	TEI	GND, GeoNames	Personen, Orte, Werke, Schriften (Typen, Schriftarten, Formen)	RDF	OWL und tw. FRBR, SKOS	9 [°]
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2013	2037	nein	GND	Briefe, Personen, Körperschaften	nein	<i>Nein</i>	1
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2018	2023	TEI	NDB, GeoNames	<i>Personen, Werke, Institutionen, Orte</i>	<i>nein</i>	<i>nein</i>	1[°]

Haussknecht	Die Reisen des Botanikers Carl Haussknecht (1838 - 1903) in das Osmanische Reich und nach Persien (1865 und 1866 - 1869). Die kommentierte digitale Edition seiner Tagebücher	2017	2020	TEI	GND, GeoNames, PlantList	Personen, Orte, Schlagworte, Pflanzen	nein	7
	Die wissenschaftliche Korrespondenz des Historikers Karl Hegel (1813-1901). Vollständige und ausschließlich digitale Edition, Einleitung, Kommentar und Register	2017	2021	TEI	GND, OSM-Geokoordinate n	Personen, Orte, Sachbegriffe, Quellen, Literaturtitel	nein	1
Hirt	Aloys Hirt – Briefwechsel 1787–1837	2014	2017/2020	TEI	GND, GeoNames	Personen, Organisationen/Institutionen, Orte, Werke, Artefakte, Schlagworte	nein	7
Humboldt (Edition)	edition Humboldt digital	2015	2032	TEI	GND/VIAF, GeoNames, Taxonomie-Datenbanken für Pflanzennamen	Personen, Institutionen, Orte, Primär- und Sekundärliteraturangaben, Pflanzennamen	nein	1
	Hidden Kosmos — Reconstructing ...	2014	2016	TEI	GND/VIAF, Wikidata	Personen, stellenweise Werke	nein	1
Humboldt & Ehrenberg	Alexander von Humboldt - Christian Gottfried Ehrenberg – Briefwechsel	XXX	XXX	XXX	XXX	XXX	XXX	X
	Online-Edition der Predigten und Briefe von Johann Friedrich Wilhelm Jerusalem (1709 - 1789)	XXX	XXX	XXX	XXX	XXX	XXX	X
Johnson	Uwe Johnson-Werkausgabe	2014	2038	TEI	GND/VIAF, LCCN	Personen, Orte, Werke, Institutionen u.a.	geplant	2 (geplant)
Jonas Cohn	Qualitative Digitalisierung des handschriftlichen Nachlasses des Philosophen und Pädagogen Jonas Cohn (1869-1947)	XXX	XXX	XXX	XXX	XXX	XXX	X
	Die Kabinettsprotokolle der Landesregierung von Nordrhein-Westfalen	2011 [digital]	ff.	TEI	nein	Personen, Orte, Organisationen, Datumsangaben, Gesetze, Verordnungen und Drucksachen, Besoldungsgruppen, „Schwärmungen“	nein	---
Karstadt	Kritische Gesamtausgabe der Schriften und Briefe Andreas Bodensteins von Karstadt	XXX	XXX	XXX	XXX	XXX	XXX	X
Kessler	Hybrid-Edition des Tagebuches von Harry Graf Kessler	XXX	XXX	XXX	XXX	XXX	XXX	X
Kommerell	Kommentierte Edition der Briefe von Max Kommerell (1902-1944)	XXX	XXX	XXX	XXX	XXX	XXX	X
Königreich Ungarn	Hungarus Digitalis - Digitale Quellenedition - Königreich Ungarn	XXX	XXX	XXX	XXX	XXX	XXX	X
Kundiges Buch	Das Kundige Bok	XXX	XXX	XXX	XXX	XXX	XXX	X
Lavater	Lavater: Historisch-kritischen Edition ausgewählter Briefwechsel (JCLB)	2017	2020/2026	Überführung in TEI 5 vorgesehen	GND	Personen, Orte, Werke, Periodika, Bibelstellen, Zitate, Worterklärungen, Schlagworte/Themen	RDF-Basis geplant (über NIE-INE, DaSCH)	8 (geplant)
							Nicht intern, aber Ontologien bei NIE-INE nach CIDOC CRM und FRBR modelliert (geplant)	

Lehndorff	Lebenswelten, Erfahrungsräume und politische Horizonte der ostpreussischen Adelsfamilie Lehndorff vom 18. bis in das 20. Jahrhundert	2016	2019	TEI	GND, GeoNames	Personen, Orte, Institutionen, Schlagworte	nein	nein	1
Mächttekongresse	Mächttekongresse 1818-1822. Digitale Edition der Dokumente zu den Kongressen von Aachen (1818), Troppau (1820), Laibach (1821) und Verona (1822)	2017	2019	TEI	VIAF, GeoNames	Personen, Orte, Daten, Verträge	nein	nein	1
Marchfutterurbar	Digitale Edition des steirisch-landesfürstlichen Marchfutterurbars von 1414/1426	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Mendelssohn Bartholdy	Historisch-kritische Online-Edition der Korrespondenz Felix Mendelssohn Bartholdys	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Mendelssohn, Erich & Luise	EMA - Der Briefwechsel von Erich und Luise Mendelssohn 1910-1953	2011	2014	ja	GND, LCAuth, GeoNames	Personen, Orte, Institutionen, Periodika, Werke, hist. Ereignisse	XXX	XXX	1
Mühsam	Erich Mühsam - Tagebücher	2011	n.n.	nein	nein	nein	nein	nein	---
Museum Fridericianum	Das Museum Fridericianum als ein Ziel von Bildungs- und Forschungsreisen der europäischen Aufklärung	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Neologie	Bibliothek der Neologie	2014	2026	TEI	GND, CERL	Personen, Orte, Werke	RDF	Ja, bei der Modellierung bibliographischer Ressourcen (CIDOC)	3
Okopenko	Andreas Okopenko: Tagebücher aus dem Nachlass (Hybridedition)	2015	2018	TEI	GND u.a.	Personen, institutionen, Orte, Werke	nein	nein	1
Pacelli	Eugenio Pacelli - Kritische Online-Edition der Nuntiaturberichte von 1917-1929	2008	2019	projektiertes XML, später Überführung in TEI	GND/VIAF	Personen, Schlagworte (darin enthalten auch Publikationen, Zeitungen und Zeitschriften etc.)	geplant	nein	2 (geplant)
Paul	Jean Paul – Sämtliche Briefe digital	2017	2018	TEI	GND, GeoNames	Personen, Orte, Werke	nein	Nein	1
Regensburger Reichstag	Der Regensburger Reichstag des Jahres 1576. Ein Pilotprojekt zum digitalen Edieren frühneuzeitlicher Quellen	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Rostocker Matrikelportal	Rostocker Matrikelportal - Datenbankedition der Immatrikulationen an der Universität Rostock 1419-1945	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
Sandrat	Joachim von Sandrat: Teutsche Akademie der Bau-, Bild- und Mahlerey-Künste ...	2006/2008	2012	TEI	GND (damals noch PND), Getty Union List of Artist Names, VIAF, Getty Thesaurus of Geographic Names, GeoNames	Personen, Orte, (Kunst-)Werke, Literatur	RDF	Eigenes Modell, das „als CIDOC CRM beeinflusst bezeichnet werden darf“	3

Sauer & Seuffert	Briefwechsel Saucher-Seuffert
Schatullenrechnungen	Die Schatullenrechnungen Friedrichs des Großen
Schlegel	Digitale Edition der Korrespondenz August Wilhelm Schlegels
Schleiermacher	Schleiermacher in Berlin 1808-1834. Briefwechsel, Tageskalender, Vorlesungen
Schnitzler & AutorInnen	Arthur Schnitzler – Briefwechsel mit Autorinnen und Autoren
Schnitzler (Werke)	Arthur Schnitzler: Digitale historisch-kritische Edition (Werke 1905–1931)
Senckenberg	Johann Christian Senckenberg: Digitale Edition der Tagebücher
Sozialische Briefwechsel	Zwischen Theologie, frühmoderner Naturwissenschaft und politischer Korrespondenz: Die sozialianischen Briefwechsel
Sturm	Der Sturm – Digitale Quellenedition zur Geschichte der internationalen Avantgarde
Theologen	Theologenbriefwechsel im Südwesten des Reichs in der Frühen Neuzeit (1550-1620)
Toller	Kommentierte Edition der Briefe Ernst Tollers
Wahlkapitulationen	Die Wahlkapitulationen der römisch-deutschen Könige von 1519 bis 1792 (Wahlkapitulationen und Kollegialschreiben)
Weber	Carl-Maria-von-Weber-Gesamtausgabe
Wedekind	Edition der Korrespondenz Frank Wedekinds als Online-Volltextdatenbank
Werner	Digitalisierung und elektronische Edition der Korrespondenz Abraham Gottlob Werners
Welscher Gast	Welscher Gast
Zimmermann	Bernd Alois Zimmermann - Gesamtausgabe

2015	<i>n.n.</i>	<i>TEI</i>	GND, GeoNames	Personen, Orte	<i>nein</i>	<i>nein</i>	1
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2012	<i>n.n.</i>	<i>TEI</i>	GND	<i>Namen, Körperschaften, Werke, Periodika, Orte</i>	<i>nein</i>	<i>Nein</i>	1
2012	2025	TEI	GND, GeoNames wird angestrebt	Personen, Orte, Werke, Bibelstellen	<i>nein</i>	<i>nein</i>	1
2018	2021	TEI	GND, Wikidata, GeoNames	Personen, Orte, Werke, Organisationen	geplant	<i>nein</i>	2 (geplant)
2012	2029	TEI	<i>nein</i> , aber URIs	<i>nein</i>	<i>nein</i>	<i>Nein</i>	1
2010	2015	TEI	GND	Personen, Orte	<i>nein</i>	<i>nein</i>	1
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2017	<i>n.n.</i>	TEI	GND, Geonames, Wikidata	Personen, Orte, Werke	Geplant	Modellierung der Sturm-Domände in CIDOC-CRM ist in Planung	3 (geplant)
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2012	2016	<i>nein</i>	<i>nein</i>	<i>nein</i>	<i>nein</i>	<i>Nein</i>	---
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
1993/1996	2026	TEI, MEI	GND/VIAF, GeoNames, RISM	Personen, Orte, Werke, Rollennamen	JSON LD	<i>nein</i>	2
2011 (Phase 1), 2018 (Phase 2)	2021	TEI	<i>nein</i>	Personen, Orte, Werke, Ereignisse	<i>nein</i>	<i>nein</i>	---
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X
2011	2023	TEI	GND	v.a. Personen, Orte, Schlagworte	Geplant (wahrscheinlich RDF)	Ja (CIDOC CRM)	9 ["]
XXX	XXX	XXX	XXX	XXX	XXX	XXX	X



Name: Meyer Vorname: Tik

Matr.Nr.:

Eidesstattliche Erklärung zur

- ☐ Hausarbeit *
- ☐ Bachelorarbeit *
- ☒ Masterarbeit *
- ☐ Magisterarbeit *

* Die eingereichte PDF-Datei ist mit den Printexemplaren identisch.

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel

Datenmodellierung für digitale Editionen -
Stand und Perspektiven zwischen XML/TEI, Linked
Open Data und Ontologien

um eine von mir erstmalig, selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung (ZSP-HU) geahndet werden.

Datum 18.5.2019

Unterschrift